OPTIMAL ERROR ESTIMATES
OF GALERKIN FINITE ELEMENT METHODS
FOR STOCHASTIC PARTIAL DIFFERENTIAL EQUATIONS
WITH MULTIPLICATIVE NOISE

RAPHAEL KRUSE*

ABSTRACT. We consider Galerkin finite element methods for semilinear stochastic partial differential equations (SPDEs) with multiplicative noise and Lipschitz continuous nonlinearities. We analyze the strong error of convergence for spatially semidiscrete approximations as well as a spatio-temporal discretization which is based on a linear implicit Euler-Maruyama method. In both cases we obtain optimal error estimates.

The proofs are based on sharp integral versions of well-known error estimates for the corresponding deterministic linear homogeneous equation together with optimal regularity results for the mild solution of the SPDE. The results hold for different Galerkin methods such as the standard finite element method or spectral Galerkin approximations.

1. INTRODUCTION

This article is devoted to the analysis of numerical schemes for the discretization of stochastic partial differential equations (SPDEs) with multiplicative noise. For the last 15 years this has been an very active field of research. An extensive list of references can be found in the review article [18].

Here we apply the well-established theory of Galerkin finite element methods from [27] and, in combination with optimal spatial and temporal regularity results [20, 22], we derive optimal error estimates for spatially semidiscrete as well as for spatio-temporal approximation schemes. Our analysis is suitable to treat different Galerkin methods such as the finite element method or spectral Galerkin methods in a unified setting.

We begin with a probability space $(\Omega, \mathcal{F}, P)$ together with a normal filtration $(\mathcal{F}_t)_{t \in [0,T]} \subset \mathcal{F}$. By $W : [0, T] \times \Omega \to U$ we denote an adapted Q-Wiener process with values in a separable Hilbert space $(U, (\cdot, \cdot)_U, \|\cdot\|_U)$. The covariance operator $Q : U \to U$ is assumed to be linear, bounded, self-adjoint and positive semidefinite.

Further, let $(H, (\cdot, \cdot), \|\cdot\|)$ be another separable Hilbert space and $A : D(A) \subset H \to H$ a linear operator, which is densely defined, self-adjoint, positive definite, not necessarily bounded but with compact inverse. Hence, there exists an increasing sequence of real numbers $(\lambda_n)_{n \geq 1}$ and an orthonormal basis of eigenvectors $(e_n)_{n \geq 1}$

in H such that $Ae_n = \lambda_n e_n$ and

$$0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n (\to \infty).$$

The domain of A is characterized by

$$D(A) = \Big\{ x \in H \; : \; \sum_{n=1}^{\infty} \lambda_n^2 (x, e_n)^2 < \infty \Big\}.$$

Thus, $-A$ is the generator of an analytic semigroup of contractions which is denoted by $E(t) = \mathrm{e}^{-At}$.

As our main example we have the following in mind: H is the space $L^2(\mathcal{D})$, where $\mathcal{D} \subset \mathbb{R}^d$ is a bounded domain with smooth boundary $\partial\mathcal{D}$ or a convex domain with polygonal boundary. Then, for example, let $-A$ be the Laplacian with homogeneous Dirichlet boundary conditions.

Next, we introduce the stochastic process, which we want to approximate. Let $X \colon [0, T] \times \Omega \to H$, $T > 0$, denote the mild solution [7, Ch. 7] to the stochastic partial differential equation

$$(1.1) \qquad \begin{aligned} \mathrm{d}X(t) + [AX(t) + f(X(t))] \, \mathrm{d}t &= g(X(t)) \, \mathrm{d}W(t), \text{ for } 0 \leq t \leq T, \\ X(0) &= X_0. \end{aligned}$$

Here f and g denote nonlinear operators which are Lipschitz continuous in an appropriate sense. In Section 2 we give a precise formulation of our conditions on f, g and X_0, which are also sufficient for the existence and uniqueness of X.

By definition the mild solution satisfies

$$(1.2) \qquad X(t) = E(t)X_0 - \int_0^t E(t-\sigma) f(X(\sigma)) \, \mathrm{d}\sigma + \int_0^t E(t-\sigma) g(X(\sigma)) \, \mathrm{d}W(\sigma)$$

for all $0 \leq t \leq T$.

Our aim is to analyze numerical schemes which are used to approximate the solution X. For an implementation one needs to discretize the time interval $[0, T]$ as well as the Hilbert spaces H and U, since both are potentially high or infinite dimensional.

In this paper we deal with the discretization of the time interval $[0, T]$ and the Hilbert space H. A fully discrete approximation of the mild solution X, which also includes the discretization of the space U, will be done in a forthcoming paper.

Our first result is concerned with a so called spatially semidiscrete approximation of (1.2), that is, we only discretize with respect to the Hilbert space H.

By $(S_h)_{h \in (0,1]}$ we denote a family of finite dimensional subspaces of H consisting of spatially regular functions. In our example with $H = L^2(\mathcal{D})$, S_h may be a standard finite element space or the linear span of finitely many eigenfunctions of A (see Examples 3.3 and 3.4).

Let the stochastic process $X_h \colon [0, T] \times \Omega \to S_h$ solve the stochastic evolution equation

$$(1.3) \qquad \begin{aligned} \mathrm{d}X_h(t) + [A_h X_h(t) + P_h f(X_h(t))] \, \mathrm{d}t &= P_h g(X_h(t)) \, \mathrm{d}W(t), \text{ for } 0 \leq t \leq T, \\ X_h(0) &= P_h X_0, \end{aligned}$$

where P_h denotes the (generalized) orthogonal projector onto S_h and $A_h \colon S_h \to S_h$ is a discrete version of the operator A which will be defined in (3.4).

As for the continuous problem (1.1) there exists a unique mild solution X_h to equation (1.3) which satisfies

$$(1.4) \quad X_h(t) = E_h(t)P_h X_0 - \int_0^t E_h(t-\sigma)P_h f(X_h(\sigma)) \, \mathrm{d}\sigma$$

$$+ \int_0^t E_h(t-\sigma)P_h g(X_h(\sigma)) \, \mathrm{d}W(\sigma) \text{ for } 0 \le t \le T.$$

This paper deals with the strong error of convergence. Here, strong convergence is understood in the sense of convergence with respect to the norm

$$\|Z\|_{L^p(\Omega;H)} = \left(\mathbf{E}\big[\|Z\|^p\big] \right)^{\frac{1}{p}}, \ p \ge 2,$$

where $\mathbf{E}$ is the expectation with respect to P. Therefore, strong convergence indicates a good pathwise approximation.

In many applications the aim is to approximate the expectation $\mathbf{E}[\varphi(X(T))]$, where φ is a smooth observable. This leads to the concept of weak convergence which, in the context of SPDEs, is considered in [9, 8, 11, 15]. However, as it was shown by Giles [12, 13], the strong order of convergence is also essential for developing efficient multilevel Monte Carlo methods for applications where the weak approximation is of interest.

Before we formulate our first main result let us explain two of the most crucial parameters. First, we have the parameter $r \in [0, 1)$ which controls the spatial regularity of the mild solution X. On the other hand, the parameter $h \in (0, 1]$ governs the granularity of the spatial approximation. In our example with $H = L^2(\mathcal{D})$ and under the given assumptions, the mild solution $X(t)$ maps into the fractional Sobolev space $H_0^1(\mathcal{D}) \cap H^{1+r}(\mathcal{D})$ and h denotes the maximal length of an edge in a partition of $\mathcal{D}$ into simplices.

Theorem 1.1. *Under the assumptions of Section 2 with $r \in [0, 1)$, $p \in [2, \infty)$, and Assumption 3.1 there exists a constant C, independent of $h \in (0, 1]$, such that*

$$\|X_h(t) - X(t)\|_{L^p(\Omega;H)} \le Ch^{1+r}, \text{ for all } t \in (0, T],$$

where X_h and X denote the mild solutions to (1.3) and (1.1), respectively.

Therefore, in our example of a standard finite element semidiscretization, the approximation X_h converges with order $1 + r$ to the mild solution X. Since this rate coincides with the spatial regularity of X it is called optimal (see [27, Ch. 1]).

We stress that, to the best of our knowledge, in all articles which deal with the numerical approximation of semilinear SPDEs the obtained order of convergence is of the suboptimal form $1 + r - \epsilon$ for any $\epsilon > 0$ (see [29] or [14], where also stronger Lipschitz assumptions have been imposed on f, g) or the error estimates contain a logarithmic term of the form $\log(t/h)$ as in [21].

Next, we consider a spatio-temporal discretization of the stochastic partial differential equation (1.1). Let $k \in (0, 1]$ denote a fixed time step which defines a time grid $t_j = jk$, $j = 0, 1, \dots, N_k$, with $N_k k \le T < (N_k + 1)k$.

Further, by X_h^j we denote the approximation of the mild solution X to (1.2) at time t_j. A combination of the Galerkin methods together with a linear implicit

Euler-Maruyama scheme results in the recursion

(1.5)
$$X_h^j - X_h^{j-1} + k\big(A_h X_h^j + P_h f(X_h^{j-1})\big) = P_h g(X_h^{j-1})\Delta W^j, \quad \text{for } j = 1, \dots, N_k,$$
$$X_h^0 = P_h X_0,$$

with the Wiener increments $\Delta W^j := W(t_j) - W(t_{j-1})$ which are $\mathcal{F}_{t_j}$-adapted, U-valued random variables. Consequently, X_h^j is an $\mathcal{F}_{t_j}$-adapted random variable which takes values in S_h.

Our second main result is the analogue of Theorem 1.1.

Theorem 1.2. *Under the assumptions of Section 2 with $r \in [0, 1)$, $p \in [2, \infty)$, and Assumptions 3.1 and 3.2 there exists a constant C, independent of $k, h \in (0, 1]$, such that*

$$\|X_h^j - X(t_j)\|_{L^p(\Omega;H)} \le C(h^{1+r} + k^{\frac{1}{2}}), \quad \text{for all } j = 1, \dots, N_k,$$

where X denotes the mild solution (1.2) to (1.1) and X_h^j is given by (1.5).

As above, we obtain the optimal order of convergence with respect to the spatio-temporal discretization. Since we only use the information of the driving Wiener process which is provided by the increments ΔW^j, it is a well-known fact [5] that the maximum order of convergence of the implicit Euler scheme is $\frac{1}{2}$. It is possible to overcome this barrier if one considers a Milstein-like scheme as discussed in the recent paper [19].

In our error analysis we use the results from [27, Ch. 2 and 3] which are stated under the assumption that $-A$ is the Laplace operator with homogeneous Dirichlet boundary conditions. But all proofs and techniques also hold in our more general framework of a self-adjoint, positive definite operator A. Further, we use generic constants which may vary at each appearance but are always independent of the discretization parameters h and k.

The structure of this paper is as follows: In the next section we introduce some additional notations and formulate the assumptions on f, g and X_0 which will be sufficient for the existence of the unique mild solution X to (1.1) as well as for the proofs of Theorems 1.1 and 1.2. In Section 3 we give a short review of Galerkin finite element methods and we also introduce two additional assumptions on the choice of the family of subspaces $(S_h)_{h \in (0,1]}$. As already mentioned, Section 3 also contains two concrete examples of a spatial discretization.

In Section 4 we present several lemmas which play a crucial role in the proofs of our main results. All lemmas are concerned with the spatially semidiscrete and fully discrete approximation of the deterministic homogeneous equation (4.1). We prove extensions of well-known convergence results from [27] to non-smooth initial data as well as sharp integral versions.

While Sections 5 and 6 are devoted to the proofs of Theorems 1.1 and 1.2, respectively, the final section revisits the special case of SPDEs with additive noise.

2. PRELIMINARIES

In order to formulate our assumptions on f, g and X_0 we introduce the notion of fractional powers of the linear operator A in the same way as in [26, 22]. After fixing some additional notation we give a precise formulation of our assumptions

and cite the result on the existence of a unique mild solution to (1.1) from [20] and a corresponding regularity result from [22].

For any $r \in \mathbb{R}$ the operator $A^{\frac{r}{2}} : D(A^{\frac{r}{2}}) \to H$ is given by

$$A^{\frac{r}{2}}x = \sum_{n=1}^{\infty} \lambda_n^{\frac{r}{2}} x_n e_n$$

for all

$$x \in D(A^{\frac{r}{2}}) = \left\{ x = \sum_{n=1}^{\infty} x_n e_n \; : \; (x_n)_{n \geq 1} \subset \mathbb{R} \text{ with } \|x\|_r^2 := \sum_{n=1}^{\infty} \lambda_n^r x_n^2 < \infty \right\}.$$

By defining $\dot{H}^r := D(A^{\frac{r}{2}})$ together with the norm $\|\cdot\|_r$ for $r \in \mathbb{R}$, $\dot{H}^r$ becomes a Hilbert space. Note that we have $\|x\|_r = \|A^{\frac{r}{2}}x\|$ for all $r \in \mathbb{R}$ and $x \in \dot{H}^r$.

As usual [7, 25] we introduce the separable Hilbert space $U_0 := Q^{\frac{1}{2}}(U)$ with the inner product $(u_0, v_0)_{U_0} := (Q^{-\frac{1}{2}}u_0, Q^{-\frac{1}{2}}v_0)_U$. Here $Q^{-\frac{1}{2}}$ denotes the pseudoinverse. Further, the set L_2^0 denotes the space of all Hilbert-Schmidt operators $\Phi : U_0 \to H$ with norm

$$\|\Phi\|_{L_2^0}^2 := \sum_{m=1}^{\infty} \|\Phi\psi_m\|^2,$$

where $(\psi_m)_{m \geq 1}$ is an arbitrary orthonormal basis of U_0 (for details see, for example, Proposition 2.3.4 in [25]). We also introduce the subset $L_{2,r}^0 \subset L_2^0$, $r \geq 0$, which is the subspace of all Hilbert-Schmidt operators $\Phi : U_0 \to \dot{H}^r$ together with the norm $\|\Phi\|_{L_{2,r}^0} := \|A^{\frac{r}{2}}\Phi\|_{L_2^0}$.

Let $r \in [0,1]$, $p \in [2, \infty)$ be given. As in [20, 22, 26] we impose the following conditions on f, g and X_0.

Assumption 2.1. *The nonlinear operator f maps H into $\dot{H}^{-1+r}$ and there exists a constant C such that*

$$\|f(x) - f(y)\|_{-1+r} \leq C\|x - y\| \quad \text{for all } x, y \in H.$$

Assumption 2.2. *The nonlinear operator g maps H into L_2^0 and there exists a constant C such that*

$$(2.1) \qquad \|g(x) - g(y)\|_{L_2^0} \leq C\|x - y\| \quad \forall x, y \in H.$$

Furthermore, we have that $g(\dot{H}^r) \subset L_{2,r}^0$ and

$$(2.2) \qquad \|g(x)\|_{L_{2,r}^0} \leq C(1 + \|x\|_r) \quad \text{for all } x \in \dot{H}^r.$$

Assumption 2.3. *The initial value $X_0 : \Omega \to \dot{H}^{r+1}$ is an $\mathcal{F}_0$-measurable random variable with $\mathbf{E}\left[\|X_0\|_{r+1}^p\right] < \infty$.*

Under these conditions, by [20, Theorem 1], there exists an up to modification unique mild solution $X : [0, T] \times \Omega \to H$ to (1.1) of the form (1.2). Furthermore, by the regularity results in [22], it holds true that for all $s \in [0, r+1]$ with $r \in [0, 1)$, we have

$$(2.3) \qquad \sup_{t \in [0,T]} \mathbf{E}\left[\|X(t)\|_s^p\right] < \infty$$

and there exists a constant C such that

$$(2.4) \qquad \left(\mathbf{E}\left[\|X(t_1) - X(t_2)\|_s^p\right]\right)^{\frac{1}{p}} \leq C|t_1 - t_2|^{\min(\frac{1}{2}, \frac{r+1-s}{2})}$$

for all $t_1, t_2 \in [0, T]$.

Remarks. 1.) Of course, Assumption 2.1 is satisfied under the usual condition that the mapping f is Lipschitz continuous from H to H. The reason for our slightly weaker assumption is that under this condition the order of the spatial discretization error will numerically behave in the same way for both integrals, the Lebesgue integral which contains f and the stochastic integral which contains g.

In addition, Assumption 2.1 applies to partial differential equations where a fractional power of the operator A is situated in front of a Lipschitz continuous mapping $\tilde{f} \colon H \to H$ as, for example, in the Cahn-Hilliard equation.

2.) Assumption 2.3 can be relaxed to $X_0 \colon \Omega \to H$ being an $\mathcal{F}_0$-measurable random variable with $\mathbf{E}\left[\|X_0\|^p\right] < \infty$. But, as with deterministic PDEs, this will lead to a singularity at $t = 0$ in the error estimates.

We complete this section by collecting useful facts on the semigroup $E(t)$. The smoothing property (2.5) and Lemma 2.4 *(ii)* are classical results and proofs can be found in [24]. A proof for the remaining assertions is given in [22].

Lemma 2.4. *For the analytic semigroup $E(t)$ the following properties hold true:*
  (i) For any $\nu \geq 0$ there exists a constant $C = C(\nu)$ such that

(2.5) $$\|A^\nu E(t)\| \leq C t^{-\nu} \text{ for all } t > 0.$$

  (ii) For any $0 \leq \nu \leq 1$ there exists a constant $C = C(\nu)$ such that

$$\|A^{-\nu}(E(t) - I)\| \leq C t^\nu \text{ for all } t \geq 0.$$

  (iii) For any $0 \leq \nu \leq 1$ there exists a constant $C = C(\nu)$ such that

$$\int_{\tau_1}^{\tau_2} \left\|A^{\frac{\nu}{2}} E(\tau_2 - \sigma)x\right\|^2 \mathrm{d}\sigma \leq C(\tau_2 - \tau_1)^{1-\nu} \|x\|^2 \text{ for all } x \in H, \, 0 \leq \tau_1 < \tau_2.$$

  (iv) For any $0 \leq \nu \leq 1$ there exists a constant $C = C(\nu)$ such that

$$\left\|A^\nu \int_{\tau_1}^{\tau_2} E(\tau_2 - \sigma)x \, \mathrm{d}\sigma\right\| \leq C(\tau_2 - \tau_1)^{1-\nu} \|x\| \text{ for all } x \in H, \, 0 \leq \tau_1 < \tau_2.$$

3. A short review of Galerkin finite element methods

In this section we first review the the Galerkin finite element methods used for the discretization of the Hilbert space H. Following [27, Ch. 2 and 3] we recall the definition of several discrete operators which are connected to a sequence of finite dimensional subspaces of $\dot{H}^1$. We close this section with two concrete examples, namely the standard finite element method and a spectral Galerkin method.

Let $(S_h)_{h \in (0,1]}$ be a sequence of finite dimensional subspaces of $\dot{H}^1$ and denote by $R_h \colon \dot{H}^1 \to S_h$ the orthogonal projector (or Ritz projector) onto S_h with respect to the inner product

$$a(x, y) := \left(A^{\frac{1}{2}}x, A^{\frac{1}{2}}y\right), \quad \text{for } x, y \in \dot{H}^1.$$

Thus, we have

$$a(R_h x, y_h) = a(x, y_h) \text{ for all } x \in \dot{H}^1, \, y_h \in S_h.$$

Throughout this paper we assume that the spaces $(S_h)_{h \in (0,1]}$, satisfy the following approximation property. Below we present two examples of A, H and spaces $(S_h)_{h \in (0,1]}$ which fulfill this assumption.

Assumption 3.1. *Let a sequence $(S_h)_{h \in (0,1]}$ of finite dimensional subspaces of $\dot{H}^1$ be given such that there exists a constant C with*

$$(3.1) \qquad \|R_h x - x\| \leq C h^s \|x\|_s \text{ for all } x \in \dot{H}^s, \ s \in \{1, 2\}, \ h \in (0, 1].$$

Remark. Following [23, Ch. 5.2] or [27, Ch. 1] consider the linear (elliptic) problem to find $x \in D(A) = \dot{H}^2$ such that $Ax = z$ holds for a given $z \in H$. The *weak* or *variational* formulation of this problem is: Find $x \in \dot{H}^1$ which satisfies

$$(3.2) \qquad a(x, y) = (z, y) \quad \text{for all } y \in \dot{H}^1.$$

For a given sequence of finite dimensional subspaces $(S_h)_{h \in (0,1]}$ the Galerkin approximation $x_h \in S_h$ of the weak solution x is given by the relationship

$$(3.3) \qquad a(x_h, y_h) = (z, y_h) \quad \text{for all } y_h \in S_h.$$

Note that by the representation theorem $x \in \dot{H}^1$ and $x_h \in S_h$ are uniquely determined by (3.2) and (3.3). By the definition of the Ritz projector and since (3.2) holds for all $y_h \in S_h$ we get

$$a(R_h x, y_h) = a(x, y_h) = a(x_h, y_h) \quad \text{for all } y_h \in S_h.$$

This yields $R_h x = x_h$, that is, the Ritz projection $R_h x$ coincides with the Galerkin approximation of the solution x to the elliptic problem. Hence, Assumption 3.1 is a statement about the order of convergence of the sequence $(x_h)_{h \in (0,1]}$ to x.

Next, we introduce the mapping $A_h \colon S_h \to S_h$, which is a discrete version of the operator A. For $x_h \in S_h$ we define $A_h x_h$ to be the unique element in S_h which satisfies the relationship

$$(3.4) \qquad a(x_h, y_h) = (A_h x_h, y_h) \text{ for all } y_h \in S_h.$$

Since

$$(A_h x_h, y_h) = a(x_h, y_h) = (x_h, A_h y_h) \text{ for all } x_h, y_h \in S_h,$$

as well as

$$(A_h x_h, x_h) = a(x_h, x_h) = \|x_h\|_1^2 > 0 \quad \text{for all } x_h \in S_h, \ x_h \neq 0,$$

the operator A_h is self-adjoint and positive definite on S_h. Hence, $-A_h$ is the generator of an analytic semigroup of contractions on S_h, which is denoted by $E_h(t) = \mathrm{e}^{-A_h t}$. Let $\rho \geq 0$. Similar to [27, Lemma 3.9] one shows the smoothing property

$$(3.5) \qquad \|A_h^\rho E_h(t) y_h\| \leq C t^{-\rho} \|y_h\| \quad \text{for all } t > 0,$$

where $C = C(\rho)$ is independent of $h \in (0, 1]$. Additionally, by the definition of A_h, it holds true that

$$(3.6) \qquad \|A_h^{\frac{1}{2}} y_h\|^2 = a(y_h, y_h) = \|A^{\frac{1}{2}} y_h\|^2 = \|y_h\|_1^2 \quad \text{for all } y_h \in S_h \subset \dot{H}^1.$$

Finally, let $P_h : \dot{H}^{-1} \to S_h$ be the generalized orthogonal projector onto S_h (see also [4]) defined by

$$(P_h x, y_h) = \langle x, y_h \rangle \quad \text{for all } x \in \dot{H}^{-1}, \ y_h \in S_h,$$

where $\langle \cdot, \cdot \rangle = a\big(A^{-1} \cdot, \cdot\big)$ denotes the duality pairing between $\dot{H}^{-1}$ and $\dot{H}^1$. By the representation theorem, P_h is well-defined and, when restricted to H, coincides with the standard orthogonal projector with respect to the H-inner product.

These operators are related as follows:

$$(3.7) \qquad A_h^{-1} P_h x = R_h A^{-1} x \quad \text{for all } x \in \dot{H}^{-1}$$

since

$$a(R_h A^{-1} x, y_h) = a(A^{-1} x, y_h) = \langle x, y_h \rangle = (P_h x, y_h) = a(A_h^{-1} P_h x, y_h)$$

for all $x \in \dot{H}^{-1}$, $y_h \in S_h$. Furthermore, it holds that

$$(3.8) \qquad \begin{aligned} \|A_h^{-\frac{1}{2}} P_h x\| &= \sup_{z_h \in S_h} \frac{\left|(A_h^{-\frac{1}{2}} P_h x, z_h)\right|}{\|z_h\|} = \sup_{z_h \in S_h} \frac{\left|(P_h x, A_h^{-\frac{1}{2}} z_h)\right|}{\|z_h\|} \\ &= \sup_{z_h' \in S_h} \frac{\left|\langle x, z_h' \rangle\right|}{\|A_h^{\frac{1}{2}} z_h'\|} \leq \sup_{z_h' \in S_h} \frac{\|x\|_{-1} \|z_h'\|_1}{\|A_h^{\frac{1}{2}} z_h'\|} = \|x\|_{-1}, \end{aligned}$$

for all $x \in \dot{H}^{-1}$, where the last equality is due to (3.6). Having established this we also prove the following consequence of (3.5)

$$(3.9) \qquad \left\| E_h(t) P_h x \right\| = \left\| A_h^{\frac{1}{2}} E_h(t) A_h^{-\frac{1}{2}} P_h x \right\| \leq C t^{-\frac{1}{2}} \left\| A_h^{-\frac{1}{2}} P_h x \right\| \leq C t^{-\frac{1}{2}} \left\| x \right\|_{-1}$$

for all $x \in \dot{H}^{-1}$, $t > 0$ and $h \in (0, 1]$.

The following assumption, which is concerned with the stability of the projector P_h with respect to the norm $\| \cdot \|_1$, will mainly be needed for the proof of Lemma 4.4 *(ii)* below and, consequently, also for the proof of Theorem 1.2.

Assumption 3.2. *Let a family $(S_h)_{h \in (0,1]}$ of finite dimensional subspaces of $\dot{H}^1$ be given such that there exists a constant C with*

$$(3.10) \qquad \|P_h x\|_1 \leq C \|x\|_1 \quad \text{for all } x \in \dot{H}^1,\ h \in (0, 1].$$

We conclude this section with two examples which satisfy Assumptions 3.1 and 3.2.

Example 3.3 (Standard finite element method)**.** Assume that $H = L^2(\mathcal{D})$, where $\mathcal{D} \subset \mathbb{R}^d$, $d = 1, 2, 3$, is a bounded, convex domain (a polygon if $d = 2$ or a polyhedron if $d = 3$). Let the operator A be given by $Au = -\nabla \cdot (a(x)\nabla u) + c(x)u$ with $c(x) \geq 0$ and $a(x) \geq a_0 > 0$ for $x \in D$ with Dirichlet boundary conditions. In this case it is well-known (for example, [23, Theorem 6.4] and [27, Ch. 3]) that $\dot{H}^1 = H_0^1(\mathcal{D})$ and $\dot{H}^2 = H^2(\mathcal{D}) \cap H_0^1(\mathcal{D})$, where $H^k(\mathcal{D})$, $k \geq 0$, denotes the Sobolev space of order k and $H_0^1(\mathcal{D})$ consists of all functions in $H^1(\mathcal{D})$ which are zero on the boundary. Furthermore, the norms in $H^k(\mathcal{D})$ and $\dot{H}^k$ are equivalent in $\dot{H}^k$ for $k = 1, 2$ (see [27, Lemma 3.1]).

Let $(\mathcal{T}_h)_{h \in (0,1]}$ denote a regular family of partitions of $\mathcal{D}$ into simplices, where h is the maximal meshsize. We define S_h to be the space of all continuous functions $y_h \colon \bar{\mathcal{D}} \to \mathbb{R}$, which are piecewise linear on $\mathcal{T}_h$ and zero on the boundary $\partial \mathcal{D}$. Then $S_h \subset \dot{H}^1$ and Assumption 3.1 holds by [27, Lemma 1.1] or [1, Theorem 5.4.8].

Further, if the family $(\mathcal{T}_h)_{h \in (0,1]}$ is quasi-uniform then also Assumption 3.2 is satisfied. But for a more detailed discussion of Assumption 3.2 in the context of the finite element method we refer to [2, 3, 6].

Example 3.4 (Spectral Galerkin method)**.** In the same situation as in Example 3.3 we further assume that $\mathcal{D} = (0, 1) \subset \mathbb{R}$ and $-A$ is the Laplace operator with

homogeneous Dirichlet boundary conditions. In this situation the orthonormal eigenfunctions and eigenvalues of the Laplace operator are explicitly known to be

$$\lambda_k = k^2 \pi^2 \text{ and } e_k(y) = \sqrt{2} \sin(k\pi y) \text{ for all } k \in \mathbb{N}, k \geq 1, y \in \mathcal{D}.$$

For $N \in \mathbb{N}$ set $h := \lambda_{N+1}^{-\frac{1}{2}}$ and define

$$S_h := \operatorname{span}\{e_k : k = 1, \ldots, N\}.$$

Note that $S_h \subset \dot{H}^r$ for every $r \in \mathbb{R}$. For $x \in \dot{H}^1$ we represent the Ritz projection $R_h x \in S_h$ in terms of the basis $(e_k)_{k=1}^N$. This yields $R_h x = \sum_{k=1}^N x_k^h e_k$, where the coefficients $(x_k^h)_{k=1}^N$ are given by

$$x_k^h = (R_h x, e_k) = \frac{1}{\lambda_k}(R_h x, A e_k) = \frac{1}{\lambda_k} a(R_h x, e_k) = \frac{1}{\lambda_k} a(x, e_k) = (x, e_k).$$

Hence, in this example, the Ritz projector R_h is the restriction of the orthogonal L^2-projector P_h to $\dot{H}^1$. Moreover, we have

$$\left\| (I - R_h)x \right\|^2 = \left\| (I - P_h)x \right\|^2 = \sum_{k=N+1}^{\infty} (x, e_k)^2 = \sum_{k=N+1}^{\infty} \lambda_k^{-\rho}(x, A^{\frac{\rho}{2}} e_k)^2$$

$$\leq \lambda_{N+1}^{-\rho} \sum_{k=N+1}^{\infty} (A^{\frac{\rho}{2}} x, e_k)^2 = h^{2\rho} \|x\|_\rho^2 \text{ for all } x \in \dot{H}^\rho, \ \rho = 1, 2,$$

since $\lambda_k^{-1} \leq \lambda_{N+1}^{-1} = h^2$ for all $k \geq N + 1$. Therefore, Assumption 3.1 is satisfied for the spectral Galerkin method.

That Assumption 3.2 holds is easily seen by

$$\left\| P_h x \right\|_1^2 = \left\| A^{\frac{1}{2}} \sum_{k=1}^N (x, e_k) e_k \right\|^2 = \sum_{k=1}^N \left(A^{\frac{1}{2}} x, e_k \right)^2 \leq \|x\|_1^2 \quad \text{for all } x \in \dot{H}^1.$$

A detailed presentation of spectral Galerkin methods is found in [17].

4. Error estimates of Galerkin methods for deterministic equations

This section extends error estimates from [27] for the discretization of the deterministic linear homogeneous equation

$$(4.1) \qquad \frac{\mathrm{d}}{\mathrm{d}t} u(t) + A u(t) = 0, \quad t > 0, \quad \text{with } u(0) = x,$$

to non-smooth initial data $x \in \dot{H}^{-1}$. We will also present suitable integral version of these estimates which are crucial for the derivation of the optimal error estimates.

4.1. Error estimates for a spatially semidiscrete approximation.

The following two lemmas provide some useful estimates of the operator F_h, which is given by $F_h(t) := E_h(t) P_h - E(t)$, $t \geq 0$. Note that $F_h(t)x$ can be seen as the error at time $t \geq 0$ between the weak solution u to (4.1) and u_h which solves the spatially semidiscrete equation

$$\frac{\mathrm{d}}{\mathrm{d}t} u_h(t) + A_h u_h(t) = 0, \quad t > 0, \quad \text{with } u_h(0) = P_h x,$$

for $x \in \dot{H}^{-1}$.

Lemma 4.1. *Under the Assumption 3.1 the following estimates hold for the error operator F_h.*

    (i) Let $0 \leq \nu \leq \mu \leq 2$. Then there exists a constant C such that

$$\|F_h(t)x\| \leq Ch^\mu t^{-\frac{\mu-\nu}{2}}\|x\|_\nu \text{ for all } x \in \dot{H}^\nu,\ t > 0,\ h \in (0,1].$$

    (ii) Let $0 \leq \rho \leq 1$. Then there exists a constant C such that

$$\|F_h(t)x\| \leq Ct^{-\frac{\rho}{2}}\|x\|_{-\rho} \text{ for all } x \in \dot{H}^{-\rho},\ t > 0,\ h \in (0,1].$$

    (iii) Let $0 \leq \rho \leq 1$. Then there exists a constant C such that

$$\|F_h(t)x\| \leq Ch^{2-\rho}t^{-1}\|x\|_{-\rho} \text{ for all } x \in \dot{H}^{-\rho},\ t > 0,\ h \in (0,1].$$

Proof. The proof of estimate *(i)* can be found in [27, Theorem 3.5].

    In order to prove *(ii)* we first note that the case $\rho = 0$ is true by *(i)*. Lemma 2.4 *(i)* yields

$$(4.2) \qquad\qquad \|E(t)x\| = \|A^{\frac{1}{2}}E(t)A^{-\frac{1}{2}}x\| \leq Ct^{-\frac{1}{2}}\|x\|_{-1}.$$

Together with (3.9) this proves

$$\|F_h(t)x\| \leq \|E_h(t)P_h x\| + \|E(t)x\| \leq Ct^{-\frac{1}{2}}\|x\|_{-1}$$

for all $x \in \dot{H}^{-1}$. This settles the case $\rho = 1$. The intermediate cases follow by the interpolation technique which is demonstrated in the proof of [27, Theorem 3.5].

    For *(iii)* the case $\rho = 0$ is again covered by *(i)*. Thus, it is enough to consider the case $\rho = 1$. First, by using (3.7), (3.5), and (3.1), we observe that

$$\begin{aligned}
\left\|F_h(t)x\right\| &= \left\|A_h E_h(t)A_h^{-1}P_h x - AE(t)A^{-1}x\right\| \\
&\leq \left\|A_h E_h(t)P_h(R_h A^{-1}x - A^{-1}x)\right\| + \left\|(A_h E_h(t)P_h - AE(t))A^{-1}x\right\| \\
&\leq Ct^{-1}\left\|(R_h - I)A^{-1}x\right\| + \left\|\frac{\mathrm{d}F_h}{\mathrm{d}t}(t)A^{-1}x\right\| \\
&\leq Ct^{-1}h\left\|A^{-1}x\right\|_1 + \left\|\frac{\mathrm{d}F_h}{\mathrm{d}t}(t)A^{-1}x\right\|.
\end{aligned}$$

Since $\|A^{-1}x\|_1 = \|x\|_{-1}$ the first term is already in the desired form. The last term is estimated by a slightly modified version of [27, Theorem 3.4], which gives

$$\left\|\frac{\mathrm{d}F_h}{\mathrm{d}t}(t)A^{-1}x\right\| \leq Cht^{-1}\left\|A^{-1}x\right\|_1.$$

This proves the case $\rho = 1$ and the intermediate cases follow by interpolation. $\qquad\square$

Lemma 4.2. *Let $0 \leq \rho \leq 1$. Under Assumption 3.1 the operator F_h satisfies the following estimates.*

    (i) There exists a constant C such that

$$\left\|\int_0^t F_h(\sigma)x\,\mathrm{d}\sigma\right\| \leq Ch^{2-\rho}\|x\|_{-\rho} \text{ for all } x \in \dot{H}^{-\rho},\ t > 0,\ h \in (0,1].$$

    (ii) There exists a constant C such that

$$\left(\int_0^t \left\|F_h(\sigma)x\right\|^2 \mathrm{d}\sigma\right)^{\frac{1}{2}} \leq Ch^{1+\rho}\|x\|_\rho \text{ for all } x \in \dot{H}^\rho,\ t > 0,\ h \in (0,1].$$

Proof. As in the proof of the previous lemma it is enough to show the estimates for $\rho = 0$ and $\rho = 1$. Then the intermediate cases follow by interpolation.

The proof of *(i)* with $\rho = 0$ is contained in the proof of [27, Theorem 3.3], where the notation

$$\tilde{e}(t) = \int_0^t F_h(\sigma) x \, \mathrm{d}\sigma$$

is used.

Here we present a proof of *(i)* with $\rho = 1$. To this end we use (3.7) and find the estimate

$$\left\| \int_0^t F_h(\sigma) x \, \mathrm{d}\sigma \right\| = \left\| \int_0^t \left(A_h E_h(\sigma) A_h^{-1} P_h - A E(\sigma) A^{-1} \right) x \, \mathrm{d}\sigma \right\|$$

$$\leq \left\| \int_0^t A_h E_h(\sigma) P_h \left(R_h - I \right) A^{-1} x \, \mathrm{d}\sigma \right\|$$

$$+ \left\| \int_0^t \left(A_h E_h(\sigma) P_h - A E(\sigma) \right) A^{-1} x \, \mathrm{d}\sigma \right\|$$

$$= \left\| \int_0^t \frac{\mathrm{d} E_h}{\mathrm{d}\sigma}(\sigma) P_h \left(R_h - I \right) A^{-1} x \, \mathrm{d}\sigma \right\| + \left\| \int_0^t \frac{\mathrm{d} F_h}{\mathrm{d}\sigma}(\sigma) A^{-1} x \, \mathrm{d}\sigma \right\|.$$

By the fundamental theorem of calculus, $\|P_h y\| \leq \|y\|$ for all $y \in H$, and Assumption 3.1 we have for the first term

$$\left\| \int_0^t \frac{\mathrm{d} E_h}{\mathrm{d}\sigma}(\sigma) P_h \left(R_h - I \right) A^{-1} x \, \mathrm{d}\sigma \right\| = \left\| \left(E_h(t) - I \right) P_h \left(R_h - I \right) A^{-1} x \right\|$$

$$\leq Ch \left\| A^{-1} x \right\|_1 = Ch \|x\|_{-1}.$$

For the second term we use Lemma 4.1 *(i)* with $\mu = \nu = 1$. This yields

$$\left\| \int_0^t \frac{\mathrm{d} F_h}{\mathrm{d}\sigma}(\sigma) A^{-1} x \, \mathrm{d}\sigma \right\| = \left\| \left(F_h(t) - F_h(0) \right) A^{-1} x \right\|$$

$$\leq \left\| F_h(t) A^{-1} x \right\| + \left\| (I - P_h) A^{-1} x \right\| \leq Ch \|x\|_{-1}.$$

In the last step we used the best approximation property of the orthogonal projector P_h, which, together with (3.1), gives

$$\left\| (P_h - I) y \right\| \leq \left\| (R_h - I) y \right\| \leq Ch \|y\|_1 \text{ for all } y \in \dot{H}^1.$$

It remains to prove *(ii)*. From [27, (2.28)] we have the inequality

$$\int_0^t \left\| F_h(\sigma) x \right\|^2 \mathrm{d}\sigma \leq \int_0^t \left\| (R_h - I) E(\sigma) x \right\|^2 \mathrm{d}\sigma.$$

In both cases, $\rho \in \{0, 1\}$, we have by (3.1)

$$\left\| (R_h - I) E(\sigma) x \right\| \leq Ch^{1+\rho} \left\| E(\sigma) x \right\|_{1+\rho} = Ch^{1+\rho} \left\| A^{\frac{1}{2}} E(\sigma) A^{\frac{\rho}{2}} x \right\|.$$

Applying Lemma 2.4 *(iii)* with $\nu = 1$ completes the proof. $\qquad\square$

4.2. **Error estimates for a fully discrete approximation.** In this subsection we consider a fully discrete approximation of the homogeneous equation (4.1). Our method of choice is a combination of the spatial Galerkin discretization together with the well-known implicit Euler scheme. As in Subsection 4.1 let a family of subspaces $(S_h)_{h \in (0,1]} \subset \dot{H}^1$ be given. The fully discrete approximation scheme is defined by the recursion

$$(4.3) \qquad U_h^j + k A_h U_h^j = U_h^{j-1}, \quad j = 1, 2, \dots \quad \text{with } U_h^0 = P_h x.$$

Here $k \in (0, 1]$ denotes a fixed time step and $U_h^j \in S_h$ denotes the approximation of $u(t_j)$ at time $t_j = jk$. A closed form representation of (4.3) is given by

$$(4.4) \qquad U_h^j = (I + kA_h)^{-j} P_h x, \quad j = 0, 1, 2, \dots .$$

In order to make the results from [27, Ch. 7] accessible and to indicate generalizations to other onestep methods onestep methods we introduce the rational function

$$R(z) = \frac{1}{1+z} \quad \text{for } z \in \mathbb{R}, z \neq 1.$$

By $R(kA_h)$ we denote the linear operator which is defined by

$$(4.5) \qquad R(kA_h)x = \sum_{m=1}^{N_h} R(k\lambda_{h,m})(x, \varphi_{h,m})\varphi_{h,m},$$

where $(\lambda_{h,m})_{m=1}^{N_h}$ are the positive eigenvalues of $A_h \colon S_h \to S_h$ with corresponding orthonormal eigenvectors $(\varphi_{h,m})_{m=1}^{N_h} \subset S_h$ and $\dim(S_h) = N_h$. With this notation (4.3) is equivalently written as

$$(4.6) \qquad U_h^j = R(kA_h)^j P_h x, \quad j = 0, 1, 2, \dots .$$

The characteristic function R of the implicit Euler scheme enjoys the following properties with $q = 1$:

$$(4.7) \qquad \begin{aligned} R(z) &= \mathrm{e}^{-z} + \mathcal{O}(z^{q+1}) \quad \text{for } z \to 0, \\ |R(z)| &< 1 \quad \text{for all } z > 0, \quad \text{and} \quad \lim_{z \to \infty} R(z) = 0. \end{aligned}$$

In the nomenclature of [27, Ch. 7] the rational function $R(z)$ is an approximation of e^{-z} with accuracy $q = 1$ and is said to be of type IV. A onestep scheme, whose characteristic rational function possesses the properties (4.7), is unconditionally stable and satisfies, for $\rho \in [0, 1]$, the discrete smoothing property

$$(4.8) \qquad \|A_h^\rho R(kA_h)^j x_h\| \leq Ct_j^{-\rho}\|x_h\| \quad \text{for all } j = 1, 2, \dots \text{ and } x_h \in S_h,$$

where the constant $C = C(\rho)$ is independent of k, h and j. For a proof of (4.8) we refer to [27, Lemma 7.3].

Further, as in the proof of [27, Theorem 7.1] it follows from (4.7) that there exists a constant C with

$$(4.9) \qquad |R(z) - \mathrm{e}^{-z}| \leq Cz^{q+1} \text{ for all } z \in [0, 1]$$

and there exists a constant $c \in (0, 1)$ with

$$(4.10) \qquad |R(z)| \leq \mathrm{e}^{-cz} \text{ for all } z \in [0, 1].$$

The rest of this subsection deals with estimates of the error between the discrete approximation U_h^j and the solution $u(t_j)$. For the error analysis in Section 6 it will be convenient to introduce an error operator

$$F_{kh}(t) := E_{kh}(t)P_h - E(t),$$

which is defined for all $t \geq 0$, where

$$(4.11) \qquad E_{kh}(t) := R(kA_h)^j, \quad \text{if } t \in [t_{j-1}, t_j) \text{ for } j = 1, 2, \dots .$$

The mapping $t \mapsto E_{kh}(t)$, and hence $t \mapsto F_{kh}(t)$, is right continuous with left limits. A simple consequence of (4.8) and (3.8) is the inequality

$$(4.12) \qquad \left\|E_{kh}(t)P_h x\right\| = \left\|A_h^{\frac{1}{2}} R(kA_h)^j A_h^{-\frac{1}{2}} P_h x\right\| \leq Ct_j^{-\frac{1}{2}}\|x\|_{-1} \leq Ct^{-\frac{1}{2}}\|x\|_{-1},$$

which holds for all $x \in \dot{H}^{-1}$, $h, k \in (0,1]$ and $t > 0$ with $t \in [t_{j-1}, t_j)$, $j = 1, 2, \ldots$.

The following lemma is the time discrete analogue of Lemma 4.1.

Lemma 4.3. *Under Assumption 3.1 the following estimates hold for the error operator F_{kh}.*

(i) Let $0 \leq \nu \leq \mu \leq 2$. Then there exists a constant C such that

$$\|F_{kh}(t)x\| \leq C(h^\mu + k^{\frac{\mu}{2}})t^{-\frac{\mu-\nu}{2}}\|x\|_\nu \text{ for all } x \in \dot{H}^\nu, \ t > 0, \ h, k \in (0,1].$$

(ii) Let $0 \leq \rho \leq 1$. Then there exists a constant C such that

$$\|F_{kh}(t)x\| \leq Ct^{-\frac{\rho}{2}}\|x\|_{-\rho} \text{ for all } x \in \dot{H}^{-\rho}, \ t > 0, \ h, k \in (0,1].$$

(iii) Let $0 \leq \rho \leq 1$. Then there exists a constant C such that

$$\|F_{kh}(t)x\| \leq C(h^{2-\rho} + k^{\frac{2-\rho}{2}})t^{-1}\|x\|_{-\rho} \text{ for all } x \in \dot{H}^{-\rho}, \ t > 0, \ h, k \in (0,1].$$

Proof. (i) Let $t > 0$ be such that $t_{j-1} \leq t < t_j$ and $x \in \dot{H}^\nu$. Then we get

$$\|F_{kh}(t)x\| \leq \|(R(kA_h)^j P_h - E(t_j))x\| + \|(E(t_j) - E(t))x\|.$$

For the second summand we have by Lemma 2.4 *(i)* and *(ii)*

$$\|(E(t_j) - E(t))x\| = \|A^{-\frac{\mu}{2}}(E(t_j - t) - I)A^{\frac{\mu-\nu}{2}}E(t)A^{\frac{\nu}{2}}x\|$$
$$\leq C(t_j - t)^{\frac{\mu}{2}}t^{-\frac{\mu-\nu}{2}}\|A^{\frac{\nu}{2}}x\| \leq Ck^{\frac{\mu}{2}}t^{-\frac{\mu-\nu}{2}}\|x\|_\nu.$$

Further, the first summand is the error between the exact solution u of (4.1) and the fully discrete scheme (4.6) at time t_j. For the case $\mu = \nu = 2$, [27, Theorem 7.8] gives the estimate

$$\|(R(kA_h)^j P_h - E(t_j))x\| \leq C(h^2 + k)\|x\|_2.$$

By the stability of the numerical scheme, that is (4.8) with $\rho = 0$, we also have the case $\mu = \nu = 0$. Hence,

$$(4.13) \qquad \|F_{kh}(t_j)x\| \leq C\|x\|,$$

and, as above, the constant C is independent of $h, k \in (0,1]$, $t_j > 0$, and x. The same interpolation technique, which is used in the proof of [27, Theorem 7.8], gives us the intermediate cases for $\mu = \nu$ and $\mu \in [0,2]$, that is

$$(4.14) \qquad \|(R(kA_h)^j P_h - E(t_j))x\| \leq C(h^\mu + k^{\frac{\mu}{2}})\|x\|_\mu.$$

On the other hand, [27, Theorem 7.7] proves the case $\nu = 0$ and $\mu = 2$. Hence, we have

$$\|(R(kA_h)^j P_h - E(t_j))x\| \leq C(h^2 + k)t_j^{-1}\|x\|,$$

where the constant C is again independent of $h, k \in (0,1]$, $t_j > 0$, and $x \in H$. An interpolation between this estimate and (4.13) shows

$$(4.15) \qquad \|(R(kA_h)^j P_h - E(t_j))x\| \leq C(h^\mu + k^{\frac{\mu}{2}})t_j^{-\frac{\mu}{2}}\|x\|$$

for all $\mu \in [0,2]$. For fixed $\mu \in [0,2]$ the proof of *(i)* is completed by an additional interpolation with respect to $\nu \in [0,\mu]$ between (4.14) and (4.15) and the fact that $t_j^{-\frac{\mu}{2}} \leq t^{-\frac{\mu}{2}}$.

The proof of *(ii)* works analogously. The case $\rho = 0$ is true by (4.13) and the case $\rho = 1$ follows by (4.2) and (4.12), since

$$\|F_{kh}(t)x\| \leq \|E_{kh}(t)P_h x\| + \|E(t)x\| \leq Ct^{-\frac{1}{2}}\|x\|_{-1}.$$

The intermediate cases follow by interpolation.

For *(iii)* the case $\rho = 0$ is already included in *(i)* with $\mu = 2$ and $\nu = 0$. Thus, it remains to show the case $\rho = 1$. For $t > 0$ with $t_{j-1} \leq t < t_j$ we have

$$\left\| F_{kh}(t)x \right\| \leq \left\| \left(R(kA_h)^j - E_h(t_j) \right) P_h x \right\| + \left\| \left(E_h(t_j)P_h - E(t_j) \right)x \right\|$$
$$+ \left\| \left(E(t_j) - E(t) \right)x \right\| =: T_1 + T_2 + T_3.$$

As in (4.5) we denote by $(\lambda_{h,m})_{m=1}^{N_h}$ the positive eigenvalues of A_h with corresponding orthonormal eigenvectors $(\varphi_{h,m})_{m=1}^{N_h} \subset S_h$. For T_1 we use the expansion of $P_h x$ in terms of $(\varphi_{h,m})_{m=1}^{N_h}$. This yields

$$T_1^2 = \left\| \sum_{m=1}^{N_h} \lambda_{h,m}^{\frac{1}{2}} \left(R(k\lambda_{h,m})^j - e^{-\lambda_{h,m}t_j} \right) \left(P_h x, \lambda_{h,m}^{-\frac{1}{2}}\varphi_{h,m} \right)\varphi_{h,m} \right\|^2$$
$$= \sum_{m=1}^{N_h} \lambda_{h,m} \left| R(k\lambda_{h,m})^j - e^{-k\lambda_{h,m}j} \right|^2 \left(A_h^{-\frac{1}{2}} P_h x, \varphi_{h,m} \right)^2.$$

First, we consider all summands with $k\lambda_{h,m} \leq 1$. As in the proof of [27, Theorem 7.1], by applying (4.9) with $q = 1$ and (4.10), we get

(4.16)

$$\left| R(k\lambda_{h,m})^j - e^{-k\lambda_{h,m}j} \right| = \left| \left(R(k\lambda_{h,m}) - e^{-k\lambda_{h,m}} \right) \sum_{i=0}^{j-1} R(k\lambda_{h,m})^{j-1-i} e^{-k\lambda_{h,m}i} \right|$$
$$\leq Cj \left(k\lambda_{h,m} \right)^2 e^{-c(j-1)k\lambda_{h,m}}.$$

Therefore, since $t_j = jk$ and $k\lambda_{h,m} \leq 1$ it holds true that

$$\lambda_{h,m} \left| R(k\lambda_{h,m})^j - e^{-k\lambda_{h,m}j} \right|^2 \leq C(jk)^{-2} k^2 \lambda_{h,m} \left(jk\lambda_{h,m} \right)^4 e^{-2cjk\lambda_{h,m}} e^{2ck\lambda_{h,m}}$$
$$\leq Ct_j^{-2}k,$$

where we also used that $\sup_{z \geq 0} z^4 e^{-2cz} < \infty$.

For all summands with $k\lambda_{h,m} > 1$ we get the estimate

$$\lambda_{h,m} \left| R(k\lambda_{h,m})^j - e^{-k\lambda_{h,m}j} \right|^2 < 2k^{-1} \left(k\lambda_{h,m} \right)^2 \left(\left| R(k\lambda_{h,m})^j \right|^2 + \left| e^{-k\lambda_{h,m}j} \right|^2 \right).$$

As it is shown in the proof of [27, Lemma 7.3], we have

(4.17) $$|R(z)| \leq \frac{1}{1+cz}, \quad \text{for all } z \geq 1, \text{ with } c > 0.$$

In fact, for the implicit Euler scheme this is immediately true with $c = 1$, but it also holds for all rational functions $R(z)$, which satisfy (4.7).

Together with $k\lambda_{h,m} > 1$ this yields

$$\left| k\lambda_{h,m} R(k\lambda_{h,m})^j \right|^2 \leq \left(\frac{k\lambda_{h,m}}{1+ck\lambda_{h,m}} \right)^2 \left(1+ck\lambda_{h,m} \right)^{-2(j-1)}$$
$$\leq \frac{1}{c^2} \left(1+c \right)^{-2(j-1)} = \frac{1}{c^2} e^{-2(j-1)\log(1+c)} \leq Cj^{-2}.$$

As above we also have

$$\left| k\lambda_{h,m} e^{-k\lambda_{h,m}j} \right|^2 \leq Cj^{-2}.$$

Therefore, also in the case $k\lambda_{h,m} > 1$, it holds that

$$\lambda_{h,m} \left| R(k\lambda_{h,m})^j - e^{-k\lambda_{h,m}j} \right|^2 \leq Ct_j^{-2}k.$$

Together with Parseval's identity and (3.8) we arrive at

$$T_1^2 \leq C t_j^{-2} k \sum_{m=1}^{\infty} \left(A_h^{-\frac{1}{2}} P_h x, \varphi_{h,m} \right)^2 = C t_j^{-2} k \left\| A_h^{-\frac{1}{2}} P_h x \right\|^2 \leq C t^{-2} k \|x\|_{-1}^2.$$

The term T_2 is covered by Lemma 4.1 *(iii)* which gives

$$T_2 = \left\| F_h(t_j) x \right\| \leq C h t_j^{-1} \|x\|_{-1} \leq C h t^{-1} \|x\|_{-1}.$$

Finally, for T_3 we apply Lemma 2.4 *(i)* with $\nu = 1$ and *(ii)* with $\nu = \frac{1}{2}$ and get

$$T_3 = \left\| A E(t) A^{-\frac{1}{2}} \left(E(t_j - t) - I \right) A^{-\frac{1}{2}} x \right\| \leq C t^{-1} (t_j - t)^{\frac{1}{2}} \|x\|_{-1} \leq C t^{-1} k^{\frac{1}{2}} \|x\|_{-1}.$$

Combining the estimates for T_1, T_2 and T_3 proves *(iii)* with $\rho = 1$. As usual, the intermediate cases follow by interpolation. $\qquad\square$

We also have an analogue of Lemma 4.2. A time discrete version of Lemma 4.4 *(ii)*, where the integral is replaced by a sum, is shown in [29].

Lemma 4.4. *Let $0 \leq \rho \leq 1$. Under Assumption 3.1 the operator F_{kh} satisfies the following estimates.*
(i) There exists a constant C such that

$$\left\| \int_0^t F_{kh}(\sigma) x \, \mathrm{d}\sigma \right\| \leq C \left(h^{2-\rho} + k^{\frac{2-\rho}{2}} \right) \|x\|_{-\rho}$$

for all $x \in \dot{H}^{-\rho}$, $t > 0$, and $h, k \in (0, 1]$.
(ii) Under the additional Assumption 3.2 there exists a constant C such that

$$\left(\int_0^t \left\| F_{kh}(\sigma) x \right\|^2 \mathrm{d}\sigma \right)^{\frac{1}{2}} \leq C \left(h^{1+\rho} + k^{\frac{1+\rho}{2}} \right) \|x\|_{\rho}$$

for all $x \in \dot{H}^{-\rho}$, $t > 0$, and $h, k \in (0, 1]$.

Proof. The proof of *(i)* uses a similar technique as the proof of Lemma 4.3 *(iii)*. First, without loss of generality, we can assume that $t = t_n$ for some $n \geq 0$. In fact, if $t_n < t < t_{n+1}$ then we have

$$\left\| \int_0^t F_{kh}(\sigma) x \, \mathrm{d}\sigma \right\| \leq \left\| \int_0^{t_n} F_{kh}(\sigma) x \, \mathrm{d}\sigma \right\| + \left\| \int_{t_n}^t F_{kh}(\sigma) x \, \mathrm{d}\sigma \right\|.$$

For the second term we get by Lemma 4.3 *(iii)*

$$\left\| \int_{t_n}^t F_{kh}(\sigma) x \, \mathrm{d}\sigma \right\| \leq \left\| \int_{t_n}^t \left(F_{kh}(\sigma) - F_{kh}(t) \right) x \, \mathrm{d}\sigma \right\| + \left\| \int_{t_n}^t F_{kh}(t) x \, \mathrm{d}\sigma \right\|$$

$$= \left\| \int_{t_n}^t \left(E(\sigma) - E(t) \right) x \, \mathrm{d}\sigma \right\| + (t - t_n) \left\| F_{kh}(t) x \right\|$$

$$\leq \left\| E(t_n) A^{\frac{\rho}{2}} \int_{t_n}^t E(\sigma - t_n) A^{-\frac{\rho}{2}} x \, \mathrm{d}\sigma \right\| + (t - t_n) \left\| A^{\frac{\rho}{2}} E(t) A^{-\frac{\rho}{2}} x \right\|$$

$$+ C(t - t_n) \left(h^{2-\rho} + k^{\frac{2-\rho}{2}} \right) t^{-1} \|x\|_{-\rho}.$$

We continue by applying Lemma 2.4 *(iv)* and *(i)* with $\nu = \frac{\rho}{2}$ and the fact that $(t - t_n) t^{-1} \leq 1$ which yields

$$\left\| \int_{t_n}^t F_{kh}(\sigma) x \, \mathrm{d}\sigma \right\| \leq C \left((t - t_n)^{1-\frac{\rho}{2}} + (t - t_n) t^{-\frac{\rho}{2}} + h^{2-\rho} + k^{\frac{2-\rho}{2}} \right) \|x\|_{-\rho}$$

$$\leq C \left(h^{2-\rho} + k^{\frac{2-\rho}{2}} \right) \|x\|_{-\rho}.$$

Next, we have

$$\left\| \int_0^{t_n} F_{kh}(\sigma) x \, d\sigma \right\| \leq \left\| \int_0^{t_n} \big(E_{kh}(\sigma) - E_h(\sigma) \big) P_h x \, d\sigma \right\|$$
$$+ \left\| \int_0^{t_n} \big(E_h(\sigma) P_h - E(\sigma) \big) x \, d\sigma \right\|.$$

For the second term Lemma 4.2 *(i)* yields the bound

$$\left\| \int_0^{t_n} \big(E_h(\sigma) P_h - E(\sigma) \big) x \, d\sigma \right\| \leq C h^{2-\rho} \|x\|_{-\rho}.$$

Thus, it is enough to show that

$$\left\| \int_0^{t_n} \big(E_{kh}(\sigma) - E_h(\sigma) \big) P_h x \, d\sigma \right\| \leq C k^{\frac{2-\rho}{2}} \|x\|_{-\rho},$$

where the constant $C = C(\rho)$ is independent of $h, k \in (0,1]$, $t > 0$, and $x \in \dot{H}^{-\rho}$.

We plug in the definition of E_{kh} and obtain

(4.18)
$$\left\| \int_0^{t_n} \big(E_{kh}(\sigma) - E_h(\sigma) \big) P_h x \, d\sigma \right\| \leq \left\| \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \big(R(kA_h)^j - E_h(t_j) \big) P_h x \, d\sigma \right\|$$
$$+ \left\| \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \big(E_h(t_j) - E_h(\sigma) \big) P_h x \, d\sigma \right\|.$$

As in (4.5) let $(\lambda_{h,m})_{m=1}^{N_h}$ be the positive eigenvalues of A_h with corresponding orthonormal eigenvectors $(\varphi_{h,m})_{m=1}^{N_h} \subset S_h$. Then, Parseval's identity yields for the first summand

$$\left\| \sum_{j=1}^n \int_{t_{j-1}}^{t_j} \big(R(kA_h)^j - E_h(t_j) \big) P_h x \, d\sigma \right\|^2$$
$$= \sum_{m=1}^{N_h} \left| k \sum_{j=1}^n \big(R(k\lambda_{h,m})^j - e^{-\lambda_{h,m} t_j} \big) \right|^2 \big(P_h x, \varphi_{h,m} \big)^2$$
$$\leq \sum_{m=1}^{N_h} \left(k \sum_{j=1}^n \lambda_{h,m}^{\frac{\rho}{2}} \big| R(k\lambda_{h,m})^j - e^{-\lambda_{h,m} t_j} \big| \right)^2 \big(A_h^{-\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2.$$

As in the proof of Lemma 4.3 *(iii)* we first study all summands with $k\lambda_{h,m} \leq 1$. In this case (4.16) gives

$$k \sum_{j=1}^n \lambda_{h,m}^{\frac{\rho}{2}} \big| R(k\lambda_{h,m})^j - e^{-\lambda_{h,m} t_j} \big| \leq C k \sum_{j=1}^n \lambda_{h,m}^{\frac{\rho}{2}} j \big(k\lambda_{h,m} \big)^2 e^{-c(j-1)k\lambda_{h,m}}$$
$$= C \lambda_{h,m}^{\frac{\rho+4}{2}} e^{ck\lambda_{h,m}} k^2 \sum_{j=1}^n j k e^{-cjk\lambda_{h,m}}$$
$$\leq C \lambda_{h,m}^{\frac{\rho+4}{2}} k \int_0^\infty (\sigma + k) e^{-c\lambda_{h,m}\sigma} \, d\sigma$$
$$\leq C \lambda_{h,m}^{\frac{\rho+4}{2}} k \Big(\frac{1}{(c\lambda_{h,m})^2} + \frac{k}{c\lambda_{h,m}} \Big) \leq C k^{\frac{2-\rho}{2}}.$$

For all summands with $k\lambda_{h,m} > 1$ we have the estimates

$$k\sum_{j=1}^{n}\lambda_{h,m}^{\frac{\rho}{2}}\big|R(k\lambda_{h,m})^j - e^{-\lambda_{h,m}t_j}\big|$$

$$< k^{\frac{2-\rho}{2}}\sum_{j=1}^{n}k\lambda_{h,m}\big(\big|R(k\lambda_{h,m})\big|^j + e^{-k\lambda_{h,m}j}\big)$$

$$\leq k^{\frac{2-\rho}{2}}\Big(\frac{k\lambda_{h,m}}{1+ck\lambda_{h,m}}\sum_{j=1}^{n}\big(1+c\big)^{-(j-1)} + k\lambda_{h,m}e^{-k\lambda_{h,m}}\sum_{j=1}^{n}e^{-(j-1)}\Big) \leq Ck^{\frac{2-\rho}{2}},$$

where we used (4.17) and $e^{-k\lambda_{h,m}(j-1)} < e^{-(j-1)}$ for $k\lambda_{h,m} > 1$. Altogether, this proves

$$(4.19) \quad \Big\|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\big(R(kA_h)^j - E_h(t_j)\big)P_h x\,\mathrm{d}\sigma\Big\|^2$$

$$\leq Ck^{2-\rho}\sum_{m=1}^{N_h}\big(A_h^{-\frac{\rho}{2}}P_h x, \varphi_{h,m}\big)^2 = Ck^{2-\rho}\big\|A_h^{-\frac{\rho}{2}}P_h x\big\|^2.$$

In order to complete the proof of *(i)* it remains to find an estimate for the second term in (4.18). By applying Parseval's identity in the same way as above we get

$$\Big\|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\big(E_h(t_j) - E_h(\sigma)\big)P_h x\,\mathrm{d}\sigma\Big\|^2$$

$$= \sum_{m=1}^{N_h}\Big|\sum_{j=1}^{n}\int_{t_{j-1}}^{t_j}\big(e^{-\lambda_{h,m}t_j} - e^{-\lambda_{h,m}\sigma}\big)\,\mathrm{d}\sigma\Big|^2\big(P_h x, \varphi_{h,m}\big)^2$$

$$= \sum_{m=1}^{N_h}\Big|\lambda_{h,m}^{\frac{\rho}{2}}\sum_{j=1}^{n}e^{-\lambda_{h,m}t_{j-1}}\int_{0}^{k}\big(e^{-k\lambda_{h,m}} - e^{-\lambda_{h,m}\sigma}\big)\,\mathrm{d}\sigma\Big|^2\big(A_h^{-\frac{\rho}{2}}P_h x, \varphi_{h,m}\big)^2.$$

Since it holds that

$$\int_{0}^{k}\big(e^{-k\lambda_{h,m}} - e^{-\lambda_{h,m}\sigma}\big)\,\mathrm{d}\sigma = ke^{-k\lambda_{h,m}} - \frac{1}{\lambda_{h,m}}\big(1 - e^{-k\lambda_{h,m}}\big)$$

we have

$$\Big|\lambda_{h,m}^{\frac{\rho}{2}}\sum_{j=1}^{n}e^{-\lambda_{h,m}t_{j-1}}\int_{0}^{k}\big(e^{-k\lambda_{h,m}} - e^{-\lambda_{h,m}\sigma}\big)\,\mathrm{d}\sigma\Big|^2$$

$$= \Big|\lambda_{h,m}^{\frac{\rho}{2}}\Big(ke^{-k\lambda_{h,m}} - \frac{1}{\lambda_{h,m}}\big(1 - e^{-k\lambda_{h,m}}\big)\Big)\sum_{j=1}^{n}e^{-k\lambda_{h,m}(j-1)}\Big|^2$$

$$= \lambda_{h,m}^{\rho-2}\big|k\lambda_{h,m}e^{-k\lambda_{h,m}} - \big(1 - e^{-k\lambda_{h,m}}\big)\big|^2\big(1 - e^{-k\lambda_{h,m}}\big)^{-2}.$$

Further, if $k\lambda_{h,m} \leq 1$ then it is true that

$$\big|k\lambda_{h,m}e^{-k\lambda_{h,m}} - \big(1 - e^{-k\lambda_{h,m}}\big)\big|^2 = e^{-2k\lambda_{h,m}}\big|e^{k\lambda_{h,m}} - 1 - k\lambda_{h,m}\big|^2$$

$$\leq Ck^4\lambda_{h,m}^4.$$

Thus, in this case we derive the estimate

$$\left| \lambda_{h,m}^{\frac{\rho}{2}} \sum_{j=1}^{n} e^{-\lambda_{h,m} t_{j-1}} \int_{0}^{k} \left(e^{-k\lambda_{h,m}} - e^{-\lambda_{h,m}\sigma} \right) d\sigma \right|^2$$

$$\leq C k^{2-\rho} \left(k\lambda_{h,m} \right)^{\rho} \frac{\lambda_{h,m}^2 k^2}{\left(1 - e^{-k\lambda_{h,m}}\right)^2} \leq C k^{2-\rho},$$

where we have used that the function $x \mapsto x(1 - e^{-x})^{-1}$ is bounded for all $x \in (0,1]$. On the other hand, if $k\lambda_{h,m} > 1$ then we have

$$\left| \lambda_{h,m}^{\frac{\rho}{2}} \sum_{j=1}^{n} e^{-\lambda_{h,m} t_{j-1}} \int_{0}^{k} \left(e^{-k\lambda_{h,m}} - e^{-\lambda_{h,m}\sigma} \right) d\sigma \right|^2$$

$$\leq 2\lambda_{h,m}^{\rho-2} \left(\left| k\lambda_{h,m} e^{-k\lambda_{h,m}} \right|^2 + \left| 1 - e^{-k\lambda_{h,m}} \right|^2 \right) \left(1 - e^{-k\lambda_{h,m}} \right)^{-2} \leq C k^{2-\rho},$$

since $\lambda_{h,m}^{\rho-2} < k^{2-\rho}$, $\sup_{x \geq 0} x e^{-x} < \infty$ and $(1 - e^{-k\lambda_{h,m}})^{-2} \leq (1 - e^{-1})^{-2}$. Altogether, this yields

$$\left\| \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left(E_h(t_j) - E_h(\sigma) \right) P_h x \, d\sigma \right\|^2 \leq C k^{2-\rho} \sum_{m=1}^{N_h} \left(A_h^{-\frac{\rho}{2}} P_h x, \varphi_{h,m} \right)^2$$

$$\leq C k^{2-\rho} \left\| A_h^{-\frac{\rho}{2}} P_h x \right\|^2,$$

which in combination with (4.19) and (3.8) completes the proof of *(i)* for $\rho \in \{0,1\}$. The intermediate cases follow again by interpolation.

As above we begin the proof of *(ii)* with the remark that without loss of generality we can assume that $t = t_n$ for some $n \geq 0$. In a similar way as in the proof of *(i)* we have

$$\left(\int_{t_n}^{t} \left\| F_{kh}(\sigma) x \right\|^2 d\sigma \right)^{\frac{1}{2}} \leq \left(\int_{t_n}^{t} \left\| \left(E(\sigma) - E(t) \right) x \right\|^2 d\sigma \right)^{\frac{1}{2}} + (t - t_n)^{\frac{1}{2}} \left\| F_{kh}(t) x \right\|.$$

For the second term Lemma 4.3 *(i)* with $\mu = 1 + \rho$ and $\nu = \rho$ together with $(t - t_n)t^{-1} \leq 1$ gives the desired estimate. The first summand is estimated by Lemma 2.4 *(ii)* which gives

$$\left(\int_{t_n}^{t} \left\| \left(E(\sigma) - E(t) \right) x \right\|^2 d\sigma \right)^{\frac{1}{2}} = \left(\int_{t_n}^{t} \left\| E(\sigma) A^{-\frac{\rho}{2}} \left(I - E(t-\sigma) \right) A^{\frac{\rho}{2}} x \right\|^2 d\sigma \right)^{\frac{1}{2}}$$

$$\leq \left(\int_{t_n}^{t} (t-\sigma)^{\rho} d\sigma \right)^{\frac{1}{2}} \|x\|_{\rho} \leq C k^{\frac{1+\rho}{2}} \|x\|_{\rho}.$$

Further, we have

$$\left(\int_{0}^{t_n} \left\| F_{kh}(\sigma) x \right\|^2 d\sigma \right)^{\frac{1}{2}} \leq \left(\int_{0}^{t_n} \left\| \left(E_{kh}(\sigma) - E_h(\sigma) \right) P_h x \right\|^2 d\sigma \right)^{\frac{1}{2}}$$

$$+ \left(\int_{0}^{t_n} \left\| \left(E_h(\sigma) P_h - E(\sigma) \right) x \right\|^2 d\sigma \right)^{\frac{1}{2}}$$

and Lemma 4.2 *(ii)* yields an estimate for the second summand of the form

$$\left(\int_{0}^{t_n} \left\| \left(E_h(\sigma) P_h - E(\sigma) \right) x \right\|^2 d\sigma \right)^{\frac{1}{2}} \leq C h^{1+\rho} \|x\|_{\rho}.$$

Thus it remains to show

$$\left(\int_{0}^{t_n} \left\| \left(E_{kh}(\sigma) - E_h(\sigma) \right) P_h x \right\|^2 d\sigma \right)^{\frac{1}{2}} \leq C k^{\frac{1+\rho}{2}} \|x\|_{\rho}.$$

As above, we prove this estimate for $\rho \in \{0, 1\}$. The intermediate cases follow again by interpolation. By the definition of E_{kh} we obtain the analogue of (4.18), namely

$$
(4.20) \quad
\begin{aligned}
\Big(\int_0^{t_n} & \big\| \big(E_{kh}(\sigma) - E_h(\sigma) \big) P_h x \big\|^2 \, \mathrm{d}\sigma \Big)^{\frac{1}{2}} \\
&\leq \Big(\sum_{j=1}^n \int_{t_{j-1}}^{t_j} \big\| \big(R(kA_h)^j - E_h(t_j) \big) P_h x \big\|^2 \, \mathrm{d}\sigma \Big)^{\frac{1}{2}} \\
&\quad + \Big(\sum_{j=1}^n \int_{t_{j-1}}^{t_j} \big\| \big(E_h(t_j) - E_h(\sigma) \big) P_h x \big\|^2 \, \mathrm{d}\sigma \Big)^{\frac{1}{2}}.
\end{aligned}
$$

For the square of the first summand we again apply Parseval's identity with respect to the orthonormal eigenbasis $(\varphi_{h,m})_{m=1}^{N_h} \subset S_h$ of A_h with corresponding eigenvalues $(\lambda_{h,m})_{m=1}^{N_h}$ and get

$$
\begin{aligned}
\sum_{j=1}^n \int_{t_{j-1}}^{t_j} & \big\| \big(R(kA_h)^j - E_h(t_j) \big) P_h x \big\|^2 \, \mathrm{d}\sigma \\
&= \sum_{j=1}^n k \sum_{m=1}^{N_h} \lambda_{h,m}^{-\rho} \big| R(k\lambda_{h,m})^j - \mathrm{e}^{-k\lambda_{h,m}j} \big|^2 \big(A_{h,m}^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2.
\end{aligned}
$$

For all summands with $k\lambda_{h,m} \leq 1$ we apply (4.16). This yields

$$
(4.21) \quad
\begin{aligned}
\sum_{j=1}^n \int_{t_{j-1}}^{t_j} & \big\| \big(R(kA_h)^j - E_h(t_j) \big) P_h x \big\|^2 \, \mathrm{d}\sigma \\
&\leq C \sum_{m=1}^{N_h} k\lambda_{h,m}^{-\rho} \sum_{j=1}^n j^2 \big(k\lambda_{h,m} \big)^4 \mathrm{e}^{-2c(j-1)k\lambda_{h,m}} \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 \\
&\leq C \sum_{m=1}^{N_h} k^2 \lambda_{h,m}^{4-\rho} \mathrm{e}^{2ck\lambda_{h,m}} \int_0^\infty (\sigma + k)^2 \mathrm{e}^{-2c\lambda_{h,m}\sigma} \, \mathrm{d}\sigma \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 \\
&\leq C \sum_{m=1}^{N_h} k^2 \lambda_{h,m}^{4-\rho} \Big(\frac{2}{(2c\lambda_{h,m})^3} + \frac{2k}{(2c\lambda_{h,m})^2} + \frac{k^2}{2c\lambda_{h,m}} \Big) \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 \\
&\leq C k^{1+\rho} \big\| A_h^{\frac{\rho}{2}} P_h x \big\|^2.
\end{aligned}
$$

For the remaining summands with $k\lambda_{h,m} > 1$ we use (4.17) and get the estimate

$$
\sum_{j=1}^n \big| R(k\lambda_{h,m})^j - \mathrm{e}^{-k\lambda_{h,m}j} \big|^2 \leq 2 \sum_{j=1}^n \big((1+c)^{-2j} + \mathrm{e}^{-2j} \big) \leq C,
$$

where the bound C is independent of n. Hence, also in this case we have

$$
(4.22) \quad
\begin{aligned}
\sum_{j=1}^n \int_{t_{j-1}}^{t_j} & \big\| \big(R(kA_h)^j - E_h(t_j) \big) P_h x \big\|^2 \, \mathrm{d}\sigma \\
&\leq C \sum_{m=1}^{N_h} k\lambda_{h,m}^{-\rho} \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 < C k^{1+\rho} \big\| A_h^{\frac{\rho}{2}} P_h x \big\|^2.
\end{aligned}
$$

Next, we prove a similar result for the square of the second summand in (4.20). As in the proof of part *(i)* an application of Parseval's identity yields

$$\sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left\| \big(E_h(t_j) - E_h(\sigma) \big) P_h x \right\|^2 \mathrm{d}\sigma$$

$$= \sum_{j=1}^{n} \sum_{m=1}^{N_h} \lambda_{h,m}^{-\rho} \int_{t_{j-1}}^{t_j} \left| \mathrm{e}^{-k\lambda_{h,m} j} - \mathrm{e}^{-\lambda_{h,m}\sigma} \right|^2 \mathrm{d}\sigma \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2.$$

By using the fact

$$\int_{t_{j-1}}^{t_j} \left| \mathrm{e}^{-k\lambda_{h,m} j} - \mathrm{e}^{-\lambda_{h,m}\sigma} \right|^2 \mathrm{d}\sigma \le \mathrm{e}^{-2k\lambda_{h,m}(j-1)} k \big(1 - \mathrm{e}^{-k\lambda_{h,m}} \big)^2$$

we obtain

$$\sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left\| \big(E_h(t_j) - E_h(\sigma) \big) P_h x \right\|^2 \mathrm{d}\sigma$$

$$\le \sum_{m=1}^{N_h} \lambda_{h,m}^{-\rho} k \big(1 - \mathrm{e}^{-k\lambda_{h,m}} \big)^2 \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 \sum_{j=1}^{n} \mathrm{e}^{-2k\lambda_{h,m}(j-1)}$$

$$\le \sum_{m=1}^{N_h} \lambda_{h,m}^{-\rho} k \big(1 - \mathrm{e}^{-k\lambda_{h,m}} \big)^2 \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 \big(1 - \mathrm{e}^{-2k\lambda_{h,m}} \big)^{-1}.$$

Since $1 - \mathrm{e}^{-2k\lambda_{h,m}} = (1 + \mathrm{e}^{-k\lambda_{h,m}})(1 - \mathrm{e}^{-k\lambda_{h,m}})$ we conclude

$$\sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left\| \big(E_h(t_j) - E_h(\sigma) \big) P_h x \right\|^2 \mathrm{d}\sigma$$

$$\le \sum_{m=1}^{N_h} \lambda_{h,m}^{-\rho} k \big(1 - \mathrm{e}^{-k\lambda_{h,m}} \big) \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2 \big(1 + \mathrm{e}^{-k\lambda_{h,m}} \big)^{-1}$$

$$\le \frac{1}{2} k \sum_{m=1}^{N_h} \lambda_{h,m}^{-\rho} \big(1 - \mathrm{e}^{-k\lambda_{h,m}} \big) \big(A_h^{\frac{\rho}{2}} P_h x, \varphi_{h,m} \big)^2$$

If $\rho = 0$ this simplifies to $\frac{1}{2} k \|P_h x\|^2$ since $1 - \mathrm{e}^{-k\lambda_{h,m}} \le 1$. If $\rho = 1$ then we use $1 - \mathrm{e}^{-k\lambda_{h,m}} \le k\lambda_{h,m}$ and the right hand side is bounded by $\frac{1}{2} k^2 \| A_h^{\frac{1}{2}} P_h x \|^2$. Altogether this proves

$$(4.23) \qquad \sum_{j=1}^{n} \int_{t_{j-1}}^{t_j} \left\| \big(E_h(t_j) - E_h(\sigma) \big) P_h x \right\|^2 \mathrm{d}\sigma \le C k^{1+\rho} \big\| A_h^{\frac{1}{2}} P_h x \big\|^2.$$

Finally, a combination of (4.20) with (4.21), (4.22) and (4.23) gives

$$(4.24) \qquad \left(\int_0^{t_n} \left\| \big(E_{kh}(\sigma) - E_h(\sigma) \big) P_h x \right\|^2 \mathrm{d}\sigma \right)^{\frac{1}{2}} \le C k^{\frac{1+\rho}{2}} \big\| A_h^{\frac{\rho}{2}} P_h x \big\|$$

which completes the proof of the case $\rho = 0$. For the case $\rho = 1$ we additionally use (3.6) and Assumption 3.2 which yields

$$\big\| A_h^{\frac{1}{2}} P_h x \big\| = \big\| P_h x \big\|_1 \le C \|x\|_1$$

for all $x \in \dot{H}^1$. This completes the proof of the Lemma.   □

5. Proof of Theorem 1.1

The aim of this section is to prove the strong convergence of the spatially semidiscrete approximation (1.4) to the mild solution (1.2) of (1.1).

The following two lemmas are useful for the proof of Theorem 1.1. The first lemma is a special version of [7, Lemma 7.2] and is needed to estimate the stochastic integrals. The second is a generalized version of Gronwall's lemma. A proof of this version can be found in [10] or in [16, Lemma 7.1.1].

Lemma 5.1. *For any $p \in [2, \infty)$, $t \in [0, T]$ and for any L_2^0-valued predictable process $\Phi(\sigma)$, $\sigma \in [0, t]$, we have*

$$\mathbf{E}\Big[\Big\| \int_0^t \Phi(\sigma)\, \mathrm{d}W(\sigma) \Big\|^p\Big] \leq C(p)\mathbf{E}\Big[\Big(\int_0^t \|\Phi(\sigma)\|_{L_2^0}^2\, \mathrm{d}\sigma \Big)^{\frac{p}{2}}\Big].$$

Here the constant $C(p)$ can be chosen to be

$$C(p) = \Big(\frac{p}{2}(p-1) \Big)^{\frac{p}{2}} \Big(\frac{p}{p-1} \Big)^{p(\frac{p}{2}-1)}.$$

Lemma 5.2. *Let the function $\varphi \colon [0, T] \to \mathbb{R}$ be nonnegative and continuous. If*

$$(5.1) \qquad \varphi(t) = C_1 + C_2 \int_0^t (t - \sigma)^{-1+\beta} \varphi(\sigma)\, \mathrm{d}\sigma$$

for some constants $C_1, C_2 \geq 0$ and $\beta > 0$ and for all $t \in (0, T]$, then there exists a constant $C = C(C_2, T, \beta)$ such that

$$\varphi(t) \leq CC_1, \ \textit{for all } t \in (0, T].$$

After these preparations we are ready to prove the first main theorem.

Proof of Theorem 1.1. For $t \in (0, T]$ we have by (1.2) and (1.4)

(5.2)
$$\big\| X_h(t) - X(t) \big\|_{L^p(\Omega;H)} \leq \big\| F_h(t)X_0 \big\|_{L^p(\Omega;H)}$$
$$+ \Big\| \int_0^t E_h(t-\sigma)P_h f(X_h(\sigma))\, \mathrm{d}\sigma - \int_0^t E(t-\sigma)f(X(\sigma))\, \mathrm{d}\sigma \Big\|_{L^p(\Omega;H)}$$
$$+ \Big\| \int_0^t E_h(t-\sigma)P_h g(X_h(\sigma))\, \mathrm{d}W(\sigma) - \int_0^t E(t-\sigma)g(X(\sigma))\, \mathrm{d}W(\sigma) \Big\|_{L^p(\Omega;H)},$$

where $F_h(t) = E_h(t)P_h - E(t)$. The first term is estimated by Lemma 4.1 *(i)* with $\mu = \nu = 1 + r$, which yields

$$(5.3) \qquad \big\| F_h(t)X_0 \big\|_{L^p(\Omega;H)} \leq Ch^{1+r} \big\| A^{\frac{1+r}{2}} X_0 \big\|_{L^p(\Omega;H)}.$$

The second term in (5.2) is dominated by three additional terms as follows

$$\Big\| \int_0^t E_h(t-\sigma)P_h f(X_h(\sigma))\, \mathrm{d}\sigma - \int_0^t E(t-\sigma)f(X(\sigma))\, \mathrm{d}\sigma \Big\|_{L^p(\Omega;H)}$$
$$\leq \Big\| \int_0^t E_h(t-\sigma)P_h\big(f(X_h(\sigma)) - f(X(\sigma))\big)\, \mathrm{d}\sigma \Big\|_{L^p(\Omega;H)}$$
$$+ \Big\| \int_0^t \big(E_h(t-\sigma)P_h - E(t-\sigma)\big)\big(f(X(\sigma)) - f(X(t))\big)\, \mathrm{d}\sigma \Big\|_{L^p(\Omega;H)}$$
$$+ \Big\| \int_0^t \big(E_h(t-\sigma)P_h - E(t-\sigma)\big)f(X(t))\, \mathrm{d}\sigma \Big\|_{L^p(\Omega;H)}$$
$$=: I_1 + I_2 + I_3.$$

We estimate each term separately. First note that, by interpolation between (3.5) and (3.9), we have $\|E_h(t)P_h x\| \le Ct^{-\frac{1-r}{2}}\|x\|_{-1+r}$. Together with Assumption 2.1 this yields

(5.4)
$$
\begin{aligned}
I_1 &\le \int_0^t \left\| E_h(t-\sigma)P_h\big(f(X_h(\sigma)) - f(X(\sigma))\big) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma \\
&\le C \int_0^t (t-\sigma)^{-\frac{1-r}{2}} \left\| X_h(\sigma) - X(\sigma) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma.
\end{aligned}
$$

The term I_2 is estimated by applying Lemma 4.1 *(iii)* with $\rho = 1 - r$, Assumption 2.1 and (2.4) with $s = 0$. Then we get

(5.5)
$$
\begin{aligned}
I_2 &\le \int_0^t \left\| F_h(t-\sigma)\big(f(X(\sigma)) - f(X(t))\big) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma \\
&\le Ch^{1+r} \int_0^t (t-\sigma)^{-1} \left\| X(\sigma) - X(t) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma \\
&\le Ch^{1+r} \int_0^t (t-\sigma)^{-1+\frac{1}{2}} \, \mathrm{d}\sigma \le Ch^{1+r}.
\end{aligned}
$$

Finally, the estimate for I_3 is a straightforward application of Lemma 4.2 *(i)* $\rho = 1 - r$. A further application of Assumption 2.1 and (2.3) gives

(5.6) $I_3 \le Ch^{1+r}\left\| f(X(t)) \right\|_{-1+r} \le Ch^{1+r}\left(1 + \sup_{\sigma \in [0,T]} \left\| X(\sigma) \right\|_{L^p(\Omega;H)}\right) \le Ch^{1+r}.$

The right hand side of this estimate is finite in view of (2.3). A combination of the estimates (5.4), (5.5), and (5.6) yields

(5.7)
$$
\begin{aligned}
&\left\| \int_0^t E_h(t-\sigma)P_h f(X_h(\sigma)) \, \mathrm{d}\sigma - \int_0^t E(t-\sigma)f(X(\sigma)) \, \mathrm{d}\sigma \right\|_{L^p(\Omega;H)} \\
&\qquad \le Ch^{1+r} + C \int_0^t (t-\sigma)^{\frac{r-1}{2}} \left\| X_h(\sigma) - X(\sigma) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma.
\end{aligned}
$$

Next, we estimate the norm of the stochastic integral in (5.2). First, we apply Lemma 5.1 and get

$$
\begin{aligned}
&\left\| \int_0^t E_h(t-\sigma)P_h g(X_h(\sigma)) \, \mathrm{d}W(\sigma) - \int_0^t E(t-\sigma)g(X(\sigma)) \, \mathrm{d}W(\sigma) \right\|_{L^p(\Omega;H)} \\
&\qquad \le C \left(\mathbf{E}\left[\left(\int_0^t \left\| E_h(t-\sigma)P_h g(X_h(\sigma)) - E(t-\sigma)g(X(\sigma)) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}}.
\end{aligned}
$$

The right hand side is a norm. Hence, the triangle inequality gives

$$
\begin{aligned}
&\left(\mathbf{E}\left[\left(\int_0^t \left\| E_h(t-\sigma)P_h g(X_h(\sigma)) - E(t-\sigma)g(X(\sigma)) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}} \\
&\quad \le \left\| \left(\int_0^t \left\| E_h(t-\sigma)P_h\big(g(X_h(\sigma)) - g(X(\sigma))\big) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})} \\
&\qquad + \left\| \left(\int_0^t \left\| F_h(t-\sigma)\big(g(X(\sigma)) - g(X(t))\big) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})} \\
&\qquad + \left\| \left(\int_0^t \left\| F_h(t-\sigma)g(X(t)) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})} \\
&\quad =: I_4 + I_5 + I_6.
\end{aligned}
$$

In a similar way as for I_1, we find an estimate for I_4 by using the stability of the operator $E_h(t)P_h$, that is, (3.5) with $\rho = 0$. Together with Assumption 2.2 we get

$$
\begin{aligned}
I_4 &\le C \Big\| \Big(\int_0^t \big\| X_h(\sigma) - X(\sigma) \big\|^2 \, d\sigma \Big)^{\frac{1}{2}} \Big\|_{L^p(\Omega;\mathbb{R})} \\
(5.8) \qquad &= C \Big(\Big\| \int_0^t \big\| X_h(\sigma) - X(\sigma) \big\|^2 \, d\sigma \Big\|_{L^{p/2}(\Omega;\mathbb{R})} \Big)^{\frac{1}{2}} \\
&\le C \Big(\int_0^t \big\| X_h(\sigma) - X(\sigma) \big\|_{L^p(\Omega;H)}^2 \, d\sigma \Big)^{\frac{1}{2}}.
\end{aligned}
$$

For the estimate of term I_5 we apply Lemma 4.1 (i) with $\mu = 1 + r$ and $\nu = 0$, which gives $\| F_h(t) \| \le C h^{1+r} t^{-\frac{1+r}{2}}$. Additionally, we use the fact that $\| LM \|_{L_2^0} \le \| L \| \| M \|_{L_2^0}$ for any bounded linear operator $L \colon H \to H$ and Hilbert-Schmidt operator $M \in L_2^0$. The estimate is completed by making use of Assumption 2.2 and the Hölder-continuity (2.4) with $s = 0$. Altogether, we derive

$$
\begin{aligned}
I_5 &\le C h^{1+r} \Big\| \Big(\int_0^t (t - \sigma)^{-1-r} \big\| X(\sigma) - X(t) \big\|^2 \, d\sigma \Big)^{\frac{1}{2}} \Big\|_{L^p(\Omega;\mathbb{R})} \\
(5.9) \qquad &\le C h^{1+r} \Big(\int_0^t (t - \sigma)^{-1-r} \big\| X(\sigma) - X(t) \big\|_{L^p(\Omega;H)}^2 \, d\sigma \Big)^{\frac{1}{2}} \\
&\le C h^{1+r} \Big(\int_0^t (t - \sigma)^{-r} \, d\sigma \Big)^{\frac{1}{2}} \le C h^{1+r}.
\end{aligned}
$$

Note that in the last step the generic constant C depends on $r \in [0, 1)$ and blows up as $r \to 1$.

Finally, for I_6, let $(\varphi_m)_{m \ge 1}$ denote an arbitrary orthonormal basis of the Hilbert space U_0. Then, by using Lemma 4.2 (ii) with $\rho = r$, Assumption 2.2 and (2.3), we get

$$
\begin{aligned}
I_6 &= \Big\| \Big(\sum_{m=1}^\infty \int_0^t \big\| F_h(t - \sigma) g(X(t)) \varphi_m \big\|^2 \, d\sigma \Big)^{\frac{1}{2}} \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\le C h^{1+r} \Big\| \Big(\sum_{m=1}^\infty \big\| A^{\frac{r}{2}} g(X(t)) \varphi_m \big\|^2 \Big)^{\frac{1}{2}} \Big\|_{L^p(\Omega;\mathbb{R})} \\
(5.10) \qquad &= C h^{1+r} \Big\| \big\| g(X(t)) \big\|_{L_{2,r}^0} \Big\|_{L^p(\Omega;\mathbb{R})} \\
&\le C h^{1+r} \Big(1 + \sup_{\sigma \in [0,T]} \big\| A^{\frac{r}{2}} X(\sigma) \big\|_{L^p(\Omega;H)} \Big) \le C h^{1+r}.
\end{aligned}
$$

In total, we have by (5.8), (5.9), and (5.10) that

$$
\begin{aligned}
(5.11) \qquad & \Big\| \int_0^t E_h(t - \sigma) P_h g(X_h(\sigma)) \, dW(\sigma) - \int_0^t E(t - \sigma) g(X(\sigma)) \, dW(\sigma) \Big\|_{L^p(\Omega;H)} \\
&\le C h^{1+r} + C \Big(\int_0^t \big\| X_h(\sigma) - X(\sigma) \big\|_{L^p(\Omega;H)}^2 \, d\sigma \Big)^{\frac{1}{2}}.
\end{aligned}
$$

Coming back to (5.2), by (5.3), (5.7), and (5.11) we conclude that

$$
\begin{aligned}
\big\| X_h(t) - X(t) \big\|_{L^p(\Omega;H)}^2 &\le C h^{2(1+r)} + C \int_0^t \big\| X_h(\sigma) - X(\sigma) \big\|_{L^p(\Omega;H)}^2 \, d\sigma \\
&\quad + C \Big(\int_0^t (t - \sigma)^{-\frac{1-r}{2}} \big\| X_h(\sigma) - X(\sigma) \big\|_{L^p(\Omega;H)} \, d\sigma \Big)^2.
\end{aligned}
$$

Finally, we note that

$$\int_0^t \big\|X_h(\sigma) - X(\sigma)\big\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma$$

$$= \int_0^t (t-\sigma)^{\frac{1-r}{2}} (t-\sigma)^{-\frac{1-r}{2}} \big\|X_h(\sigma) - X(\sigma)\big\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma$$

$$\leq T^{\frac{1-r}{2}} \int_0^t (t-\sigma)^{-\frac{1-r}{2}} \big\|X_h(\sigma) - X(\sigma)\big\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma,$$

and by Hölder's inequality

$$\int_0^t (t-\sigma)^{-\frac{1-r}{2}} \big\|X_h(\sigma) - X(\sigma)\big\|_{L^p(\Omega;H)} \, \mathrm{d}\sigma$$

$$= \int_0^t (t-\sigma)^{-\frac{1-r}{4}} (t-\sigma)^{-\frac{1-r}{4}} \big\|X_h(\sigma) - X(\sigma)\big\|_{L^p(\Omega;H)} \, \mathrm{d}\sigma$$

$$\leq \Big(\frac{2}{r+1} T^{\frac{r+1}{2}}\Big)^{\frac{1}{2}} \Big(\int_0^t (t-\sigma)^{-\frac{1-r}{2}} \big\|X_h(\sigma) - X(\sigma)\big\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma\Big)^{\frac{1}{2}}.$$

Hence, for $\varphi(t) = \big\|X_h(t) - X(t)\big\|_{L^p(\Omega;H)}^2$ we have shown that

$$\varphi(t) \leq Ch^{2(1+r)} + C\int_0^t (t-\sigma)^{\frac{r-1}{2}} \varphi(\sigma) \, \mathrm{d}\sigma$$

and Lemma 5.2 completes the proof. $\square$

6. Proof of Theorem 1.2

This section is devoted to the proof of Theorem 1.2. As in Section 5 the proof relies on a discrete version of Gronwall's Lemma. Here we use a variant from [10, Lemma 7.1].

Lemma 6.1. *Let $C_1, C_2 \geq 0$ and let $(\varphi_j)_{j=1,\dots,N_k}$ be a nonnegative sequence. If for $\beta \in (0,1]$ we have*

$$(6.1) \qquad \varphi_j \leq C_1 + C_2 k \sum_{i=1}^{j-1} t_{j-i}^{-1+\beta} \varphi_i \quad \text{for all } j = 1, \dots, N_k,$$

then there exists a constant $C = C(C_2, T, \beta)$ such that

$$\varphi_j \leq CC_1 \quad \text{for all } j = 1, \dots, N_k.$$

In particular, the constant C does not depend on k.

Proof of Theorem 1.2. In terms of the rational function $R(kA_h)$, which was introduced in (4.5), we derive the following discrete variation of constants formula for X_h^j

$$(6.2)$$

$$X_h^j = R(kA_h)^j P_h X_0 - k \sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h f(X_h^i) + \sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h g(X_h^i)\Delta W^{i+1}.$$

By applying (1.2) and (6.2) we get for the error

(6.3)
$$\|X_h^j - X(t_j)\|_{L^p(\Omega;H)} \leq \|R(kA_h)^j P_h X_0 - E(t_j) X_0\|_{L^p(\Omega;H)}$$

$$+ \left\| k \sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h f(X_h^i) - \int_0^{t_j} E(t_j - \sigma) f(X(\sigma)) \, d\sigma \right\|_{L^p(\Omega;H)}$$

$$+ \left\| \sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h g(X_h^i) \Delta W^{i+1} - \int_0^{t_j} E(t_j - \sigma) g(X(\sigma)) \, dW(\sigma) \right\|_{L^p(\Omega;H)}.$$

The first summand is the error of the fully discrete approximation scheme (4.3) for the homogeneous equation (4.1) but with the initial value being a random variable. By Assumption 2.3 we have that $X_0(\omega) \in \dot{H}^{1+r}$ for P-almost all $\omega \in \Omega$. Thus, the error estimate from [27, Theorem 7.8] yields

$$\|R(kA_h)^j P_h X_0 - E(t_j) X_0\|_{L^p(\Omega;H)} \leq C\left(h^{1+r} + k^{\frac{1+r}{2}}\right) \|A^{\frac{1+r}{2}} X_0\|_{L^p(\Omega;H)}.$$

For the other two summands we introduce an auxiliary process which is given by

(6.4)
$$X_{kh}(t) := X_h^{j-1}, \quad \text{if } t \in (t_{j-1}, t_j], \quad j = 1, \ldots, N_k,$$
$$X_{kh}(0) := P_h X_0.$$

By this definition X_{kh} is an adapted and left-continuous process and, therefore, predictable [25, p. 27]. Recalling the definition (4.11) of the family of operators $E_{kh}(t)$, $t \geq 0$, we obtain

(6.5)
$$k \sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h f(X_h^i) = \sum_{i=0}^{j-1} \int_{t_i}^{t_{i+1}} E_{kh}(t_j - \sigma) P_h f(X_{kh}(\sigma)) \, d\sigma$$
$$= \int_0^{t_j} E_{kh}(t_j - \sigma) P_h f(X_{kh}(\sigma)) \, d\sigma$$

and, analogously,

(6.6)
$$\sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h g(X_h^i) \Delta W^{i+1} = \int_0^{t_j} E_{kh}(t_j - \sigma) P_h g(X_{kh}(\sigma)) \, dW(\sigma).$$

By applying (6.5) we have the following estimate of the second summand in (6.3)

$$\left\| k \sum_{i=0}^{j-1} R(kA_h)^{j-i} P_h f(X_h^i) - \int_0^{t_j} E(t_j - \sigma) f(X(\sigma)) \, d\sigma \right\|_{L^p(\Omega;H)}$$

$$= \left\| \int_0^{t_j} \left(E_{kh}(t_j - \sigma) P_h f(X_{kh}(\sigma)) - E(t_j - \sigma) f(X(\sigma)) \right) d\sigma \right\|_{L^p(\Omega;H)}$$

$$\leq \left\| \int_0^{t_j} E_{kh}(t_j - \sigma) P_h \left(f(X_{kh}(\sigma)) - f(X(\sigma)) \right) d\sigma \right\|_{L^p(\Omega;H)}$$

$$+ \left\| \int_0^{t_j} \left(E_{kh}(t_j - \sigma) P_h - E(t_j - \sigma) \right) \left(f(X(\sigma)) - f(X(t_j)) \right) d\sigma \right\|_{L^p(\Omega;H)}$$

$$+ \left\| \int_0^{t_j} \left(E_{kh}(t_j - \sigma) P_h - E(t_j - \sigma) \right) f(X(t_j)) \, d\sigma \right\|_{L^p(\Omega;H)}$$

$$=: J_1 + J_2 + J_3.$$

Note, that the terms J_1, J_2, and J_3 are of the same structure as the terms I_1, I_2, and I_3 in the proof of Theorem 1.1. Since with Lemmas 4.3 and 4.4 we have the

time discrete analogues of Lemmas 4.1 and 4.2 at our disposal the proof follows the same path.

For the term J_1 we first note that an interpolation between (4.8) and (4.12) yields $\left\| E_{kh}(t) P_h x \right\| \leq C t_i^{-\frac{1-r}{2}} \left\| x \right\|_{-1+r}$ for all $t \in [t_{i-1}, t_i)$ and all $x \in \dot{H}^{-1+r}$. Together with Assumption 2.1 this gives

$$
\begin{aligned}
J_1 &\leq \int_0^{t_j} \left\| E_{kh}(t_j - \sigma) P_h \big(f(X_{kh}(\sigma)) - f(X(\sigma)) \big) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma \\
&\leq C \sum_{i=0}^{j-1} \int_{t_i}^{t_{i+1}} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(\sigma) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma \\
&\leq C k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)} \\
&\quad + C \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \int_{t_i}^{t_{i+1}} \left\| X(t_i) - X(\sigma) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma.
\end{aligned}
$$

By the Hölder continuity (2.4) we get

$$
\begin{aligned}
\sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} & \int_{t_i}^{t_{i+1}} \left\| X(t_i) - X(\sigma) \right\|_{L^p(\Omega;H)} \mathrm{d}\sigma \\
&\leq C \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \int_{t_i}^{t_{i+1}} (\sigma - t_i)^{\frac{1}{2}} \mathrm{d}\sigma = \frac{2}{3} C k^{\frac{3}{2}} \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \\
&\leq C k^{\frac{1}{2}} \int_0^{t_j} \sigma^{-\frac{1-r}{2}} \mathrm{d}\sigma \leq C k^{\frac{1}{2}}.
\end{aligned}
$$

Altogether, J_1 is estimated by

$$
(6.7) \qquad J_1 \leq C k^{\frac{1}{2}} + C k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}.
$$

For the estimate of J_2 we first note that $F_{kh}(t) = E_{kh}(t) P_h - E(t)$ and, hence, we can apply Lemma 4.3 *(iii)* with $\rho = 1 - r$. In the same way as in (5.5) we obtain

$$
(6.8) \qquad J_2 \leq C \big(h^{1+r} + k^{\frac{1+r}{2}} \big).
$$

Likewise, but this time by an application of Lemma 4.4 *(i)* with $\rho = 1 - r$, we proceed with the term J_3 in same way as in (5.6). By also using (2.3) this yields

$$
(6.9) \qquad J_3 \leq C \big(h^{1+r} + k^{\frac{1+r}{2}} \big) \Big(1 + \sup_{\sigma \in [0,T]} \left\| X(\sigma) \right\|_{L^p(\Omega;H)} \Big) \leq C \big(h^{1+r} + k^{\frac{1+r}{2}} \big).
$$

It remains to estimate the third summand in (6.3) which contains the stochastic integrals. With (6.6) and Lemma 5.1 we get

$$
\begin{aligned}
\Big\| \sum_{i=0}^{j-1} & R(k A_h)^{j-i} P_h g(X_h^i) \Delta W^{i+1} - \int_0^{t_j} E(t_j - \sigma) g(X(\sigma)) \,\mathrm{d}W(\sigma) \Big\|_{L^p(\Omega;H)} \\
&= \Big\| \int_0^{t_j} E_{kh}(t_j - \sigma) P_h g(X_{kh}(\sigma)) \,\mathrm{d}W(\sigma) - \int_0^{t_j} E(t_j - \sigma) g(X(\sigma)) \,\mathrm{d}W(\sigma) \Big\|_{L^p(\Omega;H)} \\
&\leq C \Big(\mathbf{E} \Big[\Big(\int_0^{t_j} \left\| E_{kh}(t_j - \sigma) P_h g(X_{kh}(\sigma)) - E(t_j - \sigma) g(X(\sigma)) \right\|_{L_2^0}^2 \mathrm{d}\sigma \Big)^{\frac{p}{2}} \Big] \Big)^{\frac{1}{p}}.
\end{aligned}
$$

In the last step Lemma 5.1 is applicable since by our definitions (4.11) and (6.4) the process $[0, t_j] \ni \sigma \mapsto E_{kh}(t_j - \sigma)P_h g(X_{kh}(\sigma)) \in L_2^0$ is adapted and left-continuous and, therefore, predictable.

Next, we use the triangle inequality and obtain

$$
\left(\mathbf{E} \left[\left(\int_0^{t_j} \left\| E_{kh}(t_j - \sigma)P_h g(X_{kh}(\sigma)) - E(t_j - \sigma)g(X(\sigma)) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}}
$$

$$
\leq \left\| \left(\int_0^{t_j} \left\| E_{kh}(t_j - \sigma)P_h \big(g(X_{kh}(\sigma)) - g(X(\sigma)) \big) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})}
$$

$$
+ \left\| \left(\int_0^{t_j} \left\| F_{kh}(t_j - \sigma) \big(g(X(\sigma)) - g(X(t_j)) \big) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})}
$$

$$
+ \left\| \left(\int_0^{t_j} \left\| F_{kh}(t_j - \sigma)g(X(t_j)) \right\|_{L_2^0}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})}
$$

$$
=: J_4 + J_5 + J_6.
$$

For the estimate of J_4 we use the facts that $\|E_{kh}(t)P_h x\| \leq C\|x\|$ for all $x \in H$ as well as $\|LM\|_{L_2^0} \leq \|L\|\|M\|_{L_2^0}$ for all linear bounded operators $L\colon H \to H$ and $M \in L_2^0$. Together with Assumption 2.2 and the same technique as in (5.8) we get

$$
J_4 \leq C \left\| \left(\int_0^{t_j} \left\| X_{kh}(\sigma) - X(\sigma) \right\|^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}} \right\|_{L^p(\Omega;\mathbb{R})}
$$

$$
\leq C \left(\int_0^{t_j} \left\| X_{kh}(\sigma) - X(\sigma) \right\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}}
$$

$$
= C \left(k \sum_{i=0}^{j-1} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2 \right)^{\frac{1}{2}}
$$

$$
+ C \left(\sum_{i=0}^{j-1} \int_{t_i}^{t_{i+1}} \left\| X(t_i) - X(\sigma) \right\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma \right)^{\frac{1}{2}}.
$$

By the Hölder continuity (2.4) it holds that

$$
\sum_{i=0}^{j-1} \int_{t_i}^{t_{i+1}} \left\| X(t_i) - X(\sigma) \right\|_{L^p(\Omega;H)}^2 \, \mathrm{d}\sigma \leq C \sum_{i=0}^{j-1} \int_{t_i}^{t_{i+1}} (\sigma - t_i) \, \mathrm{d}\sigma \leq Ck.
$$

Hence, we have

$$
(6.10) \qquad J_4 \leq Ck^{\frac{1}{2}} + C \left(k \sum_{i=0}^{j-1} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2 \right)^{\frac{1}{2}}.
$$

In the way same as in (5.9), we apply Lemma 4.3 (i) with $\mu = 1 + r$ and $\nu = 0$ and we derive the estimate

$$
(6.11) \qquad J_5 \leq C \big(h^{1+r} + k^{\frac{1+r}{2}} \big),
$$

where we also used Assumption 2.2 and (2.4). By (2.3) and the same arguments which gave (5.10) we get

$$
(6.12) \qquad J_6 \leq C \big(h^{1+r} + k^{\frac{1+r}{2}} \big).
$$

To summarize our estimate of (6.3), by (6.7) to (6.12) we have shown that

$$\left\| X_h^j - X(t_j) \right\|_{L^p(\Omega;H)}^2 \le C\big(h^{1+r} + k^{\frac{1}{2}} + k^{\frac{1+r}{2}} \big)^2$$
$$+ C\Big(k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)} \Big)^2$$
$$+ C k \sum_{i=0}^{j-1} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2.$$

Further, we have

$$k \sum_{i=0}^{j-1} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2 \le T^{\frac{1-r}{2}} k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2,$$

and by Hölder's inequality

$$k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}$$
$$\le \Big(k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \Big)^{\frac{1}{2}} \Big(k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2 \Big)^{\frac{1}{2}}$$
$$\le \Big(\frac{2}{r+1} T^{\frac{r+1}{2}} \Big)^{\frac{1}{2}} \Big(k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \left\| X_h^i - X(t_i) \right\|_{L^p(\Omega;H)}^2 \Big)^{\frac{1}{2}}.$$

Hence, by setting $\varphi_j = \left\| X_h^j - X(t_j) \right\|_{L^p(\Omega;H)}^2$ we have proven that

$$\varphi_j \le C\big(h^{1+r} + k^{\frac{1}{2}} + k^{\frac{1+r}{2}} \big)^2 + C k \sum_{i=0}^{j-1} t_{j-i}^{-\frac{1-r}{2}} \varphi_i.$$

An application of Lemma 6.1 completes the proof of the theorem. □

7. Additive noise

In this section we focus on stochastic partial differential equations with additive noise, that is, the L_2^0-valued function g does not depend on X. Thus, the SPDE (1.1) has the form

(7.1)
$$\mathrm{d}X(t) + \big[AX(t) + f(X(t)) \big] \mathrm{d}t = g \, \mathrm{d}W(t), \quad \text{for } 0 \le t \le T,$$
$$X(0) = X_0.$$

Numerical schemes for the approximation of SPDEs with additive noise have been extensively studied. For example, for schemes which involve the finite element method we refer to [14, 21, 28] and the references therein.

For additive noise Assumption 2.2 simplifies to

Assumption 7.1. *There exists $r \in [0,1]$ such that the Hilbert-Schmidt operator $g \in L_2^0$ satisfies*

$$\|g\|_{L_{2,r}^0} < \infty.$$

Recall that in the case, where $U = H$ and $g \colon H \to H$ is the identity, Assumption 7.1 reads as follows

$$\|g\|_{L_{2,r}^0} = \|A^{\frac{r}{2}} g\|_{L_2^0} = \sum_{m=1}^{\infty} \|A^{\frac{r}{2}} Q^{\frac{1}{2}} \varphi_m\|^2 < \infty,$$

where $(\varphi_m)_{m \geq 1}$ denotes an arbitrary orthonormal basis of H. In particular, if $r = 0$ we have $\|g\|_{L_2^0} = \mathrm{Tr}(Q) < \infty$ which is a common assumption on the covariance operator Q (see [7, 25]).

Under Assumptions 2.1, 2.3 and 7.1 with $r \in [0, 1]$, $p \in [2, \infty)$ there exists a mild solution $X: [0, T] \times \Omega \to \dot{H}^{1+r}$ to (7.1) (see [7] and [22, Corollary 5.2]).

In particular, we stress the fact that the parameter value $r = 1$ is included for SPDEs with additive noise. As the next corollary shows the same is true for the error estimates of the numerical approximations (1.4) and (1.5).

Corollary 7.2. *Let Assumptions 2.1, 2.3 and 7.1 hold for some $r \in [0, 1]$, $p \in [2, \infty)$. Let X denote the mild solution to (7.1).*

(i) Under Assumption 3.1 there exists a constant C, independent of $h \in (0, 1]$, such that

$$\left\| X_h(t) - X(t) \right\|_{L^p(\Omega;H)} \leq C h^{1+r}, \quad \text{for all } t \in (0, T],$$

where X_h is the corresponding spatially semidiscrete approximation (1.4).

(ii) Under Assumptions 3.1 and 3.2 there exists a constant C, independent of $k, h \in (0, 1]$, such that

$$\left\| X_h^j - X(t_j) \right\|_{L^p(\Omega;H)} \leq C(h^{1+r} + k^{\frac{1}{2}}), \quad \text{for all } j = 1, \ldots, N_k,$$

where X_h^j is the corresponding spatio-temporal discrete approximation (1.5).

Proof. For *(i)* and *(ii)* the assertion follows directly from Theorems 1.1 and 1.2 for all parameter values $r \in [0, 1)$.

A close inspection of the proof of Theorem 1.1 shows that the condition $r < 1$ is only required in the estimate (5.9). But in the case of additive noise the term I_5 is equal to 0. Hence, the convergence result holds true by the same arguments for $r = 1$.

The same is true for *(ii)* where the condition $r < 1$ only shows up in the estimate of (6.11). □

We remark that a similar convergence result to *(i)* can be found in [21, Prop. 2.3]. But there the authors had to incorporate a singularity of the form $\max(0, \log(\frac{t}{h^2}))$ which is now removed.

References

[1] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[2] C. Carstensen. Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for H^1-stability of the L^2-projection onto finite element spaces. *Math. Comp.*, 71(237):157–163 (electronic), 2002.

[3] C. Carstensen. An adaptive mesh-refining algorithm allowing for an H^1 stable L^2 projection onto Courant finite element spaces. *Constr. Approx.*, 20(4):549–564, 2004.

[4] K. Chrysafinos and L. S. Hou. Error estimates for semidiscrete finite element approximations of linear and semilinear parabolic equations under minimal regularity assumptions. *SIAM J. Numer. Anal.*, 40(1):282–306, 2002.

[5] J. M. C. Clark and R. J. Cameron. The maximum rate of convergence of discrete approximations for stochastic differential equations. In *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978)*, volume 25 of *Lecture Notes in Control and Information Sci.*, pages 162–171. Springer, Berlin, 1980.

[6] M. Crouzeix and V. Thomée. The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces. *Math. Comp.*, 48(178):521–532, 1987.

[7] G. Da Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*, volume 44 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1992.

[8] A. Debussche. Weak approximation of stochastic partial differential equations: the nonlinear case. *Math. Comp.*, 80(273):89–117, 2011.

[9] A. Debussche and J. Printems. Weak order for the discretization of the stochastic heat equation. *Math. Comp.*, 78(266):845–863, 2009.

[10] C. M. Elliott and S. Larsson. Error estimates with smooth and nonsmooth data for a finite element method for the Cahn-Hilliard equation. *Math. Comp.*, 58(198):603–630, S33–S36, 1992.

[11] M. Geissert, M. Kovács, and S. Larsson. Rate of weak convergence of the finite element method for the stochastic heat equation with additive noise. *BIT Numer. Math.*, 49:343–356, 2009.

[12] M. B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[13] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[14] E. Hausenblas. Approximation for semilinear stochastic evolution equations. *Potential Anal.*, 18(2):141–186, 2003.

[15] E. Hausenblas. Weak approximation of the stochastic wave equation. *J. Comput. Appl. Math.*, 235(1):33–58, 2010.

[16] D. Henry. *Geometric Theory of Semilinear Parabolic Equations*, volume 840 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1981.

[17] J. S. Hesthaven, S. Gottlieb, and D. Gottlieb. *Spectral Methods for Time-Dependent Problems*, volume 21 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2007.

[18] A. Jentzen and P. E. Kloeden. The numerical approximation of stochastic partial differential equations. *Milan J. Math.*, 77:205–244, 2009.

[19] A. Jentzen and M. Röckner. A break of the complexity of the numerical approximation of nonlinear SPDEs with multiplicative noise. *Preprint, arXiv:1001.2751v2*, 2010.

[20] A. Jentzen and M. Röckner. Regularity analysis for stochastic partial differential equations with nonlinear multiplicative trace class noise. *Preprint, arXiv:1005.4095v1*, 2010.

[21] M. Kovács, S. Larsson, and F. Lindgren. Strong convergence of the finite element method with truncated noise for semilinear parabolic stochastic equations with additive noise. *Numer. Algorithms*, 53(2-3):309–320, 2010.

[22] R. Kruse and S. Larsson. Optimal regularity for semilinear stochastic partial differential equations with multiplicative noise. *Preprint*, 2010. (submitted).

[23] S. Larsson and V. Thomée. *Partial Differential Equations with Numerical Methods*, volume 45 of *Texts in Applied Mathematics*. Springer-Verlag, Berlin, 2003.

[24] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*, volume 44 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1983.

[25] C. Prévôt and M. Röckner. *A Concise Course on Stochastic Partial Differential Equations*, volume 1905 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.

[26] J. Printems. On the discretization in time of parabolic stochastic partial differential equations. *M2AN Math. Model. Numer. Anal.*, 35(6):1055–1078, 2001.

[27] V. Thomée. *Galerkin Finite Element Methods for Parabolic Problems*, volume 25 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 2006.

[28] Y. Yan. Semidiscrete Galerkin approximation for a linear stochastic parabolic partial differential equation driven by an additive noise. *BIT*, 44(4):829–847, 2004.

[29] Y. Yan. Galerkin finite element methods for stochastic parabolic partial differential equations. *SIAM J. Numer. Anal.*, 43(4):1363–1384, 2005.